

MANUAL BEWERTUNG VON SYSTEMATISCHEN REVIEWS ZU THERAPEUTISCHEN INTERVENTIONEN

Version 2.0 vom 29.11.2022



Die initiale Version dieses Manuals mit dem Titel „Bewertung von systematischen Übersichtsarbeiten: ein Manual für die Leitlinienerstellung“ wurde im Rahmen des vom Bundesministerium für Gesundheit (BMG) geförderten Projekts „ZEIG“ von Cochrane Deutschland in Kooperation mit dem Institut für Medizinisches Wissensmanagement der AWMF (AWMF-IMWi) erstellt. Die vorliegende Aktualisierung dieses Manual wurde ohne externe Projektförderung erstellt.

Kommentare zu diesem Manual sind ausdrücklich erwünscht und können gerichtet werden an: [cgs@cochrane.de](mailto:cds@cochrane.de)

KONTAKTE:

¹Cochrane Deutschland Stiftung

Berliner Allee 2
D-79110 Freiburg
www.cochrane.de

²Institut für Evidenz in der Medizin

Universitätsklinikum Freiburg
Medizinische Fakultät
Breisacher Straße 86
79110 Freiburg
www.uniklinik-freiburg.de/institut-fuer-evidenz-in-der-medizin

³Ärztliches Zentrum für Qualität in der Medizin

Tiergarten Tower
Straße des 17. Juni 106-108
10623 Berlin
www.aezq.de

⁴Institut für Versorgungsforschung und Gesundheitsökonomie, Centre for Health and Society

Medizinische Fakultät und Universitätsklinik Düsseldorf
Heinrich Heine-Universität Düsseldorf
Postfach 132
Moorenstraße 5
40225 Düsseldorf
www.uniklinik-duesseldorf.de/patienten-besucher/klinikeninstitutezentren/institut-fuer-versorgungsforschung-und-gesundheitsoekonomie

⁵Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement (AWMF-IMWi)

Karl von Frisch Str. 1
Philipps Universität
D-35043 Marburg
www.awmf.org/leitlinien/awmf-imwi.html

Bitte wie folgt zitieren:

Cochrane Deutschland, Institut für Evidenz in der Medizin, Ärztliches Zentrum für Qualität in der Medizin, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement. „Manual Bewertung von systematischen Reviews zu therapeutischen Interventionen“. Version 2.0, 2022. Verfügbar bei: Cochrane Deutschland <https://www.cochrane.de/literaturbewertung>; ÄZQ: <https://www.leitlinien.de/methodik/>; AWMF: <https://www.awmf.org/regelwerk/downloads>.
DOI: 10.6094/UNIFR/231856, <https://freidok.uni-freiburg.de/data/231856>

Urheberrecht:

Das Werk ist in allen seinen Teilen urheberrechtlich geschützt. Die vorliegenden Texte dürfen für den persönlichen Gebrauch (gemäß § 53 UrhG) in einer EDV-Anlage gespeichert und (in inhaltlich unveränderter Form) ausgedruckt werden. Bitte beachten Sie, dass nur das unter <https://www.cochrane.de/literaturbewertung>, <https://www.leitlinien.de/methodik/> und <https://www.awmf.org/regelwerk/downloads> verfügbare Dokument gültig ist. Verweise ("links") aus anderen Dokumenten des World Wide Web auf das Manual unter den vorstehenden Adressen sind ohne weiteres zulässig und erwünscht, für eine entsprechende Mitteilung sind wir jedoch dankbar. Jede darüber hinausgehende, insbesondere kommerzielle, Verwertung bedarf der schriftlichen Zustimmung der angegebenen Urheber und/oder Inhabern von Verwertungsrechten.

AUTOR*INNEN:

Braun C^{1,2}, Schmucker C², Schaefer C³, Möhler R⁴, Nothacker M⁵, Meerpohl JJ^{1,2}

LAYOUT:

Georg Melber¹

¹Cochrane Deutschland Stiftung

Berliner Allee 2
D-79110 Freiburg

²Institut für Evidenz in der Medizin

Breisacher Straße 86
79110 Freiburg

³Ärztliches Zentrum für Qualität in der Medizin

Tiergarten Tower
Straße des 17. Juni 106-108
10623 Berlin

⁴Institut für Versorgungsforschung und Gesundheitsökonomie, Centre for Health and Society

Medizinische Fakultät und Universitätsklinik Düsseldorf
Postfach 132
Moorenstraße 5
40225 Düsseldorf

⁵Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement (AWMF-IMWi)

Karl von Frisch Str. 1
Philipps Universität
D-35043 Marburg

Wir bedanken uns bei Prof. Dr. Stefan Sauerland für die Durchsicht und Kommentierung der vollständigen aktualisierten Manualversion; bei Dr. Isolde Sommer für die Durchsicht und Kommentierung des Abschnitts zu Rapid Reviews (Kapitel 1.2); bei Prof. Dr. Dawid Pieper für die Durchsicht und Kommentierung des Abschnitts zu den Bewertungsinstrumenten AMSTAR 2 und ROBIS (Kapitel 3.2) und bei Dr. Gerta Rücker und Dr. Guido Schwarzer für die Durchsicht und Überarbeitung der weiterführenden Informationen zu Metaanalysen und Disseminationsbias (Anhang); und bei Dr. Birgit Schindler für die redaktionelle Durchsicht der finalen Version.

INHALTSVERZEICHNIS

ABKÜRZUNGSVERZEICHNIS	7
TABELLEN- UND ABBILDUNGSVERZEICHNIS	8
1 EINLEITUNG	9
1.1 Systematische Reviews	9
1.2 Rapid Reviews und Living Systematic Reviews	12
1.3 Andere Evidenzsynthese-Formate	12
1.4 Systematische Reviews für Leitlinien und Health Technology Assessments	13
1.5 Wesentliche Neuerungen gegenüber der initialen Version des Manuals	14
1.6 Ziele und Struktur des Manuals	15
2 ANFORDERUNGEN AN SYSTEMATISCHE REVIEWS	16
2.1 Methodische Qualität	16
2.2 Biasrisiko	16
2.3 Berichtqualität	17
2.4 Übertragbarkeit der Reviewergebnisse	17
2.5 Metaanalysen	18
2.6 Vertrauenswürdigkeit der Evidenz	20
3 DIE KRITISCHE BEWERTUNG VON SYSTEMATISCHEN REVIEWS	21
3.1 Einführung und Übersicht	21
3.2 Die Bewertungsinstrumente AMSTAR 2 und ROBIS	23
4 DIE BEWERTUNG DER METHODISCHEN QUALITÄT VON SYSTEMATISCHEN REVIEWS MIT AMSTAR 2	25
4.1 Hintergrund	25
4.2 Welche Arten von systematischen Reviews können mit AMSTAR 2 bewertet werden?	25
4.3 Wie ist AMSTAR 2 aufgebaut?	26
4.4 Wie wird die AMSTAR 2-Bewertung durchgeführt?	28

5 DIE BEWERTUNG DES BIASRISIKOS IN SYSTEMATISCHEN REVIEWS MIT ROBIS	30
5.1 Hintergrund	30
5.2 Welche Arten von systematischen Reviews können mit ROBIS bewertet werden?	31
5.3 Wie ist ROBIS aufgebaut?	31
5.4 Wie wird die ROBIS-Bewertung durchgeführt?	32
5.5 Wie wird die ROBIS-Bewertung dokumentiert und präsentiert?	34
6 STANDARDS FÜR DIE BERICHTERSTATTUNG ZU SYSTEMATISCHEN REVIEWS – PRISMA 2020	35
7 LITERATUR	37
ANHANG: WEITERFÜHRENDE INFORMATIONEN ZU METAANALYSEN UND DISSEMINATIONSBIAS	41

ABKÜRZUNGSVERZEICHNIS

AMSTAR	A Measurement Tool to Assess Systematic Reviews
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
RoB	Risk of Bias
CASP	Critical Appraisal Skills Programme
CERqual	Confidence in the Evidence from Reviews of Qualitative research
DART	Documentation Appraisal Review Tool
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HEBW	Health Evidence Bulletins - Wales: Critical Appraisal of a systematic review
HTA	Health Technology Assessment
IPD	Individual participant data
MA	Metaanalyse
MECIR	Methodological Expectations of Cochrane Intervention Reviews
NMA	Netzwerkmetaanalyse
CNMA	Komponenten-Netzwerkmetaanalyse (component network metaanalysis)
NRSI	Non-randomised study of intervention
OQAQ	Overview Quality Assessment Questionnaire
PICO	Population, Intervention, Comparison, Outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	International prospective register of systematic reviews
R-AMSTAR	Revised Assessment of Multiple SysTemAtic Reviews
RCT	Randomised Controlled Trial; randomisierte kontrollierte Studie
ROBIS	Risk Of Bias In Systematic Reviews
SIGN	Scottish Intercollegiate Guidelines Network
TRANSFER	Transferability of Systematic Review Findings

TABELLEN- UND ABBILDUNGSVERZEICHNIS

TAB. 1: BEWERTUNGSINSTRUMENTE FÜR SYSTEMATISCHE REVIEWS	22
TAB. 2: DIE 16 ITEMS/FRAGEN DER AMSTAR 2 CHECKLISTE	27
ABB. 1: AMSTAR 2 - ITEM/FRAGE 9	28
ABB. 2: AMSTAR 2 - BEWERTUNG DES GESAMT-VERTRAUENS IN DIE ERGEBNISSE DES REVIEWS	29
ABB. 3: DIE VIER ROBIS-DOMÄNEN	31
ABB. 4: ROBIS – DOMÄNE 2	32
ABB. 5: ROBIS – PHASE 3, SIGNALFRAGEN	33
ABB. 6: ROBIS - GESAMTEINSCHÄTZUNG DES BIASRISIKOS DES REVIEWS	34
ABB. 7: PRISMA 2020 STATEMENT: WESENTLICHE DOKUMENTE UND RESSOURCEN	36
ABB. 8: FUNNEL PLOT	46

1 EINLEITUNG

1.1 Systematische Reviews

Im Gesundheitswesen tätige Praktiker*innen, Wissenschaftler*innen, Leitlinienentwickler*innen und andere wissenschaftlich interessierte Personen und Gruppen teilen ein Problem: der stetige große Zuwachs an Publikationen zu neuen (Primär-) Studien macht es immer schwieriger, Schritt zu halten und auf dem aktuellen wissenschaftlichen Kenntnisstand zu gesundheitsbezogenen Fragestellungen zu bleiben. Sämtliche in einem Fachgebiet neu publizierte Studien zu lesen, ist in aller Regel unmöglich. Unabhängig hiervon ist die Betrachtung einzelner Studien zu einer Fragestellung ohne Berücksichtigung anderer (idealerweise aller) zu der Fragestellung verfügbarer Studien ohnehin problematisch, denn sie kann zu Fehleinschätzungen und falschen Schlüssen führen. Andere zu der Fragestellung verfügbare Studien könnten zu anderen, möglicherweise sogar gegensätzlichen, Ergebnissen gekommen sein. Dies kann z. B. dazu führen, dass eine therapeutische Intervention als wirksamer oder sicherer eingeschätzt wird, als sie es in Wirklichkeit ist. Für eine verlässliche Einschätzung der Ergebnisse wissenschaftlicher Studien zu einer Fragestellung bedarf es einer Übersicht über die gesamte verfügbare Evidenz.

Die genannten Schwierigkeiten im Umgang mit Primärstudien sind der wesentliche Grund für den hohen Stellenwert, den hochwertige **systematische Reviews** als kompakte, gleichzeitig aber umfassende Quelle für die Erfassung des wissenschaftlichen Kenntnisstandes zu gesundheitsbezogenen Fragestellungen haben. Ein systematischer Review ist eine wissenschaftliche Arbeit, in der die (idealerweise gesamte) verfügbare empirische Evidenz zu einer **spezifischen Fragestellung** unter Anwendung **expliziter, systematischer Methoden identifiziert, kritisch bewertet** und **zusammengefasst** ist.¹ Die Zusammenfassung der Evidenz kann dabei qualitativ oder quantitativ mit statistischen Methoden, d. h. mit Metaanalysen, erfolgen. Die Zusammenfassung der Ergebnisse mehrerer Studien durch eine Metaanalyse erhöht die Aussagekraft eines Ergebnisses gegenüber dem Ergebnis aus einer Einzelstudie,² ist jedoch nicht immer möglich oder sinnvoll.

Systematische Reviews sind von **narrativen Reviews** abzugrenzen, die keiner expliziten, systematischen Methodik folgen und damit kein verlässliches Gesamtbild der zu einer Fragestellung verfügbaren Evidenz bieten. Die Auswahl der berücksichtigten Literatur erfolgt in einem narrativen

Review in aller Regel nicht-systematisch und selektiv, was das Risiko von Subjektivität und einem erheblichen Bias der Ergebnisse und Schlussfolgerungen birgt. Ein publizierter narrativer Review ist von einem systematischen Review in der Regel vor allem dadurch unterscheidbar, dass ein strukturierter Methodenteil mit den für systematische Reviews charakteristischen Angaben (u.a. zu den Kriterien für den Ein- und Ausschluss von Studien in den Review, zur systematischen Suche nach relevanten Studien und zu den Analysemethoden) fehlt. Zudem fehlt in der Regel eine kritische Bewertung der einbezogenen Studien. Narrative Reviews können einen hilfreichen Überblick über ein Thema geben, sollten aber wegen ihrer Anfälligkeit für Bias nicht für Aussagen über die Wirksamkeit und/oder Sicherheit von Interventionen herangezogen werden.

Systematische Reviews können zu unterschiedlichen **Fragestellungen** (z.B. zu therapeutischen Interventionen, diagnostischen, prognostischen oder qualitativen Fragestellungen) und entsprechend zu unterschiedlichen **Studientypen** durchgeführt werden. **Dieses Manual fokussiert auf systematische Reviews zu therapeutischen Interventionen**, d. h. zu (randomisierten und nicht-randomisierten) vergleichenden Interventionsstudien. Die in dem Manual dargestellten Grundprinzipien der kritischen Bewertung von systematischen Reviews lassen sich jedoch auch auf andere Arten von systematischen Reviews übertragen.

Im Fall von systematischen Reviews zur Wirksamkeit und Sicherheit therapeutischer Interventionen gelten **randomisierte kontrollierte Studien (RCTs)** grundsätzlich als das am besten geeignete, am wenigsten für Bias anfällige Studiendesign. RCTs werden in der klinischen Forschung als methodischer **Goldstandard für die Evaluation der Wirksamkeit und Sicherheit (neuer) Therapien** betrachtet. Sie werden zur Untersuchung patientenrelevanter klinischer Fragestellungen durchgeführt und stellen in der Arzneimittelentwicklung eine wesentliche Grundlage für behördliche Zulassungsentscheidungen dar. Systematische Reviews zu therapeutischen Interventionen können jedoch auch andere Studientypen beinhalten, wie z. B. nicht-randomisierte Interventionsstudien (NRSI) oder einarmige Kohortenstudien. Die Auswahl der eingeschlossenen Studientypen sollte immer begründet werden. Ein Grund für den Einschluss nicht-randomisierter Interventionsstudien kann der bekannte oder antizipierte Mangel an RCTs sein, ein Grund für den Einschluss von einarmigen Kohortenstudien die Ermittlung von selten auftretenden unerwünschten Ereignissen, die in RCTs oftmals nicht oder allenfalls unzureichend erfasst werden können.

Bei anderen Fragestellungen, wie etwa zur Ätiologie oder Prognose einer Erkrankung oder zu qualitativen Erhebungen von Patientenerfahrungen, besteht die geeignete Evidenzbasis für einen systematische Review entsprechend aus anderen Studientypen.³

Systematische Reviews zu therapeutischen Interventionen können, je nach Fragestellung und Ausrichtung, ausschließlich quantitative Studientypen (z. B. RCTs, NRSI oder einarmige Kohortenstudien) beinhalten, oder durch Evidenz aus qualitativen Studien ergänzt werden. Ein systematischer Review, der Evidenz sowohl aus quantitativen als auch aus qualitativen Studien beinhaltet, wird als „**Mixed Methods Systematic Review**“ bezeichnet.⁴ Ziel des Mixed Methods-Ansatzes ist es, ein tieferes Verständnis für Aspekte wie die Komplexität von Interventionen, ihre Implementierung und/oder die Präferenzen und Erfahrungen von Patient*innen zu erlangen.

Der Fokus dieses Manuals liegt auf der kritischen Bewertung von systematischen Reviews zu quantitativen Interventionsstudien. Weiterführende Informationen zu systematischen Reviews zu qualitativer Evidenz und Mixed Methods Systematic Reviews finden sich u.a. im Cochrane Handbook for Systematic Reviews of Interventions.⁴ Für die kritische Bewertung der methodischen Qualität und Vertrauenswürdigkeit von qualitativer Evidenz gibt es eine Vielzahl von Instrumenten, z. B. die **CASP Qualitative Study Checklist**⁵ (für Studien) oder der **GRADE Confidence in the Evidence from Review of Qualitative Research (CERQual)**-Ansatz⁶ (für systematische Reviews).

Ein **Cochrane Review** ist ein systematischer Review, der nach den methodischen Standards von **Cochrane**⁷ erstellt und in der **Cochrane Library**⁸ publiziert wurde. Die Erstellung eines Cochrane Reviews wird durchgängig, beginnend mit der Registrierung des Review-Titels, von Cochrane begleitet. Vergleichende Untersuchungen haben gezeigt, dass die methodische Qualität von Cochrane Reviews im Allgemeinen höher ist als die anderer („non-Cochrane“) systematischer Reviews.⁹⁻¹¹ Die meisten Cochrane Reviews untersuchen therapeutische Interventionen, jedoch gibt es zunehmend auch Cochrane Reviews zu diagnostischen, prognostischen, qualitativen und weiteren Fragestellungen.

1.2 Rapid Reviews und Living Systematic Reviews

Ein **Rapid Review („schneller Review“)** ist, wie der Name impliziert, ein systematischer Review, der in einem vergleichsweise kurzen Zeitraum erstellt wird. Rapid Reviews sind in der Literatur bislang unterschiedlich definiert und beschrieben worden.¹² Hamel et al. schlugen 2021 die folgende vorläufige Definition vor (eigene Übersetzung):^{12(S.80)} „Ein Rapid Review ist eine Form von Wissenssynthese, bei der der Erstellungsprozess eines traditionellen systematischen Reviews durch das Verschließen oder Weglassen verschiedener methodischer Aspekte beschleunigt wird, um Evidenz auf eine ressourceneffiziente Weise zusammenzufassen.“ Methodische Anpassungen können z. B. in Beschränkungen des Umfangs der Literatursuche oder der Analysen bestehen. Rapid Reviews wurden primär entwickelt, um dringliche Fragestellungen zu bearbeiten, für die eine schnelle Entscheidungsgrundlage erforderlich ist.^{12,13} Weiterführende Informationen zu Rapid Reviews bieten u. a. die Ressourcen der **Cochrane Rapid Reviews Methods Group**.^{13,14}

Ein neuerer Ansatz für die Aktualisierung von systematischen Reviews ist der **Living Systematic Review-Ansatz** (LSR; „Lebender systematischer Review“). Ein Living Systematic Review ist ein systematischer Review, der kontinuierlich aktualisiert wird, d. h. in den relevante neue Evidenz eingeschlossen wird, sobald sie verfügbar ist.^{15,16} Dieser Ansatz wurde primär entwickelt, um einer hohen Forschungsdynamik und dem damit einhergehenden Risiko des schnellen Überholtseins wissenschaftlicher Erkenntnisse gerecht zu werden. Living Systematic Reviews gewinnen u. a. für die Aktualisierung von Leitlinien (**Living guideline recommendations**) zunehmend an Bedeutung.¹⁷

1.3 Andere Evidenzsynthese-Formate

Es gibt eine Reihe weiterer **Evidenzsynthese-Formate**, von denen wir im Folgenden einige kurz darstellen, unter anderem, um ihre wesentlichen Unterschiede zum Format des „klassischen“ systematischen Reviews darzustellen. Die in diesem Manual vorgestellten Instrumente zur kritischen Bewertung von systematischen Reviews lassen sich nur teilweise und in unterschiedlichem Ausmaß auf diese Evidenzsynthese-Formate übertragen.

Zunehmend werden auch systematische Reviews erstellt, die sich anstelle von Evidenz aus Primärstudien auf aggregierte Evidenz aus bereits existierenden systematischen Reviews beziehen. Dieses Review-Format wird meist als **Overview**, mitunter aber auch als **Meta-Review, Overview of**

Reviews oder **Umbrella Review** bezeichnet.^{18,19} Das Ziel eines Overviews ist es, ressourcensparend eine Evidenzbasis für die Gesundheitsversorgung zu schaffen.¹⁹ Häufig liegen jedoch zwischen der interessierenden Fragestellung und der Fragestellung der identifizierten aggregierten Evidenz Differenzen vor („scope mismatch“). In solchen Fällen muss dann trotzdem auf die in der aggregierten Evidenz enthaltenen primären Publikationen zurückgegriffen werden.²⁰

Zu den neueren Evidenzsynthese-Formaten zählen zudem **Evidence Maps** oder **Scoping Reviews**.^{21,22} Evidence Maps oder Scoping Reviews werden in der Regel zu dem Zweck erstellt, den Umfang und wesentliche Charakteristika der verfügbaren wissenschaftlichen Literatur zu einer breit gefassten Fragestellung systematisch und in kompakter (häufig tabellarischer) Form abzubilden. Scoping Reviews liefern zudem eine deskriptive Zusammenfassung der identifizierten Literatur. Eine kritische Bewertung der eingeschlossenen Studien erfolgt dabei nicht immer. Durch die Erfassung des Umfangs der verfügbaren wissenschaftlichen Literatur bzw. Evidenz erlauben diese Formate insbesondere auch das Aufzeigen von Forschungslücken.

1.4 Systematische Reviews für Leitlinien und Health Technology Assessments

Systematische Reviews spielen in der **Entwicklung und Aktualisierung von evidenzbasierten Leitlinien** eine bedeutende Rolle. Leitlinien sind systematisch entwickelte Aussagen, die den gegenwärtigen Wissensstand wiedergeben, um die Entscheidungsfindung von Ärzt*innen sowie Angehörigen von weiteren Gesundheitsberufen und Patient*innen/Bürger*innen für eine angemessene Versorgung bei spezifischen Gesundheitsproblemen zu unterstützen. Sie sollten auf einer systematischen Sichtung und Bewertung der Evidenz und einer Abwägung von Nutzen und Schaden alternativer Vorgehensweisen basieren.²³

Wo verfügbar, werden systematische Reviews zunehmend als Grundlage für die Zusammenfassung der Evidenz zu den in einer Leitlinie enthaltenen Fragestellungen und Empfehlungen herangezogen. Gemäß dem Regelwerk der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF-Regelwerk)²³ wird der folgende iterative, hierarchische Rechercheprozess für die Erstellung von Leitlinien empfohlen: (i) Berücksichtigung von aggregierter Evidenz (systematische Reviews ohne oder mit Metaanalysen), (ii) ggf. Berücksichtigung von Leitlinien, (iii) Berücksichtigung von Primärliteratur. Dies bedeutet, dass zunächst gezielt nach

systematischen Reviews zu suchen ist. Kann die bestverfügbare Evidenz nicht durch aktuelle, hochwertige systematischen Reviews abgebildet werden oder – fakultativ – durch aktuelle evidenzbasierte Leitlinien, werden Primärstudien herangezogen.

Systematische Reviews sind zudem eine wichtige Grundlage für **Health Technology Assessments (HTAs)**. Ein HTA ist ein multidisziplinärer Prozess, bei dem explizite Methoden angewandt werden, um den Wert einer Gesundheitstechnologie (**Health Technology**, z.B. eines diagnostischen Verfahrens, eines Medikaments, eines Medizinproduktes oder eines Versorgungsprozesses) zu unterschiedlichen Zeitpunkten ihres Lebenszyklus zu ermitteln.²⁴ Ziel dieses Prozesses ist es, Informationen für die Entscheidungsfindung als Grundlage für ein faires, effizientes und hochwertiges Gesundheitssystem bereitzustellen. Der Prozess der Erstellung eines HTA ist formal, systematisch und transparent und erfolgt unter Anwendung neuester wissenschaftlicher Methoden unter Berücksichtigung der besten verfügbaren Evidenz.²⁴ Da HTAs in der Regel auf systematischen Reviews basieren, stellen sie eine weitere hilfreiche Datengrundlage für Leitlinien dar. Für ihre kritische Bewertung können die in diesem Manual dargestellten Instrumenten verwendet werden.

1.5 Wesentliche Neuerungen gegenüber der initialen Version des Manuals

Die initiale Version dieses Manuals richtete sich explizit primär an Leitlinienerstellende und -beratende. Da die Inhalte jedoch über diese Gruppe hinaus auch für eine breitere Zielgruppe relevant und von Interesse sind, haben wir den Fokus erweitert: Diese Version des Manuals richtet sich somit an alle wissenschaftlich tätigen Personen und Gruppen, die sich mit der kritischen Bewertung von therapeutischen Interventionsstudien in systematischen Reviews befassen bzw. sich für diese interessieren.

Das Manual wurde umfassend überarbeitet und aktualisiert. Zu den wesentlichen inhaltlichen Neuerungen zählen die Aufnahme des AMSTAR 2 Tools (Kapitel 4) und der PRISMA 2020-Ressourcen (Kapitel 8). Aufbau und Struktur des Manuals wurden in Teilen verändert, das Inhaltsverzeichnis entsprechend angepasst. Alle Kapitel wurden editiert und teilweise angepasst bzw. ergänzt. Die Literatur wurde aktualisiert und ergänzt.

1.6 Ziele und Struktur des Manuals

Das vorliegende Manual soll wissenschaftlich tätige Personen und Gruppen wie Autor*innen von systematischen Reviews und HTAs sowie Leitlinienerstellende und -beratende, aber auch Entscheidungsträger im Gesundheitswesen und Personen oder Gruppen, die sich für die kritische Bewertung von Interventionsstudien interessieren, in ihrer Arbeit mit wissenschaftlicher Literatur unterstützen. Es soll Nutzenden einen Überblick über die qualifizierte kritische Bewertung von systematischen Reviews vermitteln und in Verbindung mit praktischer Anleitung (z.B. im Rahmen eines Workshops oder Leitlinienseminars) zur eigenständigen Bewertung der methodischen Qualität und des Biasrisikos von systematischen Reviews befähigen.

Die Grundlagen dieses Manuals bilden international anerkannte wissenschaftliche Standards für systematische Reviews und ihre Bewertung.⁷ Das Manual ergänzt und vertieft u. a. das **AWMF-Regelwerk Leitlinien**,²³ insbesondere den Abschnitt „Kritische Bewertung der Evidenz“.

Das Manual ist in verschiedene Kapitel gegliedert; Kernthemen sind **Anforderungen an systematische Reviews (Kapitel 2)**, **die Bewertung der methodischen Qualität von systematischen Reviews mit AMSTAR 2 (Kapitel 4)** und **die Bewertung des Biasrisikos in systematischen Reviews mit ROBIS (Kapitel 5)**. In einem weiteren Kapitel werden die **PRISMA 2020-Ressourcen für die Berichterstattung von systematischen Reviews (Kapitel 6)** vorgestellt. Im Anhang finden sich weiterführend Informationen zu **Metaanalysen und Disseminationsbias**.

2 ANFORDERUNGEN AN SYSTEMATISCHE REVIEWS

2.1 Methodische Qualität

Die methodische Qualität eines systematischen Reviews bezieht sich auf das Ausmaß, in dem er dem höchstmöglichen methodischen Standard entsprechend durchgeführt und berichtet wurde.²⁵ Grundsätzlich soll bei der Erstellung eines systematischen Reviews von der Planung über die Durchführung bis zur Auswertung und Publikation nach einem standardisierten Konzept vorgegangen werden. Dies beinhaltet die Erstellung eines Protokolls, in dem a priori beschrieben wird, wie bei seiner Erstellung vorgegangen werden soll. Anstelle eines publizierten Protokolls kann auch eine prospektive Registrierung eines systematischen Reviews, z.B. in der Datenbank **PROSPERO**²⁶, erfolgen.

Der aktuelle methodische Standard zur Durchführung von Cochrane Reviews ist den **MECIR** (**M**ethodological **E**xpectations of **C**ochrane **I**ntervention **R**eviews) **Standards**²⁷ oder dem **Cochrane Handbook for Systematic Reviews of Interventions**⁷ zu entnehmen. Die MECIR Standards beinhalten neben Details zur standardisierten Planung und Durchführung auch Empfehlungen zur Berichterstattung von systematischen Reviews zu therapeutischen Interventionen.

2.2 Biasrisiko

Der Begriff Bias (Verzerrung) bezieht sich im Kontext von systematischen Reviews auf systematische Abweichungen der Ergebnisse oder Schlussfolgerungen eines Reviews von der „Wahrheit“, zu denen es infolge von Schwächen im Design, der Durchführung, der Analysen oder der Berichterstattung kommen kann.²⁵ Bias kann insbesondere bei der Studienauswahl, der Datenextraktion, der kritischen Bewertung der eingeschlossenen Primärstudien, der Datensynthese und der Darstellung von Ergebnissen und Schlussfolgerungen auftreten. Es liegt in den Händen der Autor*innen eines systematischen Reviews, potenzielle methodische Schwächen der Primärstudien transparent darzustellen, diese bei der Ergebnisinterpretation adäquat zu berücksichtigen und eine selektive Berichterstattung zu vermeiden. Den Ergebnissen und Schlussfolgerungen eines systematischen Reviews sollte nur dann vertraut werden, wenn er ein geringes Biasrisiko aufweist und die in ihm zusammengefasste Evidenz vollständig, transparent

und nachvollziehbar abbildet. Hierbei ist zu berücksichtigen, dass auch ein hochwertiger systematischer Review mögliche Probleme der in ihm enthaltenen Primärstudien (Bias, mangelnde Präzision, eingeschränkte Übertragbarkeit) nicht lösen kann („garbage-in-garbage-out“-Problem). Für die Bewertung des Biasrisikos (**Risk of Bias, RoB**) in systematischen Reviews sind verschiedene Instrumente verfügbar (s. **Kapitel 3**).

2.3 Berichtqualität

Eine wichtige Voraussetzung für die angemessene kritische Bewertung der methodischen Qualität und des Biasrisikos eines systematischen Reviews ist eine gute **Qualität des Reviewberichts**.²⁸ Dies bedeutet, dass alle Aspekte des Reviews, die für seine Bewertung von Bedeutung sind, möglichst vollständig, ausreichend detailliert und nachvollziehbar bzw. unmissverständlich beschrieben sein sollten. Mängel in der Berichtqualität erschweren die Einschätzung und Bewertung eines systematischen Reviews. Was nicht oder nur unzureichend berichtet ist, kann nicht oder nur unzureichend bewertet werden. So bedeutet die fehlende Angabe einer unabhängigen doppelten Studienselektion und Datenextraktion nicht zwingend, dass diese nicht durchgeführt wurde, führt aber im Zweifelsfall zu einer negativen, mindestens aber zu einer unklaren Bewertung dieses wichtigen methodischen Aspekts. Eine Option bei fehlenden und unklaren Angaben ist die Kontaktaufnahme zu den Autor*innen des Reviews.

Eine **Hilfestellung insbesondere für Autor*innen für die Verbesserung der Berichterstattung zu systematischen Reviews** bieten **Standards für die Berichterstattung („reporting guidelines“)**. Weiterführende Informationen zum **PRISMA Statement**²⁹, dem bekanntesten Standard für die Berichterstattung von systematischen Reviews (zu therapeutischen Interventionen), finden sich in **Kapitel 6**.

2.4 Übertragbarkeit der Reviewergebnisse

Neben einer angemessenen Durchführung und Berichterstattung eines systematischen Reviews sollte auch der Kontext der Ergebnisse und damit ihre „**Übertragbarkeit**“ berücksichtigt werden. Die in einen systematischen Review eingeschlossenen Studien können sich erheblich darin unterscheiden, wer involviert war, wo und wann sie durchgeführt wurden und durch welche weiteren Faktoren sie beeinflusst wurden, wie z.B. die Organisation der Gesundheitsversorgung oder das politische oder soziale Umfeld.³⁰ Diese Faktoren können einen Einfluss darauf haben,

inwieweit die Ergebnisse der eingeschlossenen Studien auf den für den Review spezifizierten Kontext und den Kontext der Endnutzer (z.B. Patient*innen, Entscheidungsträger*innen) übertragbar sind.³¹ Die Berücksichtigung des Kontextes der Ergebnisse wissenschaftlicher Arbeiten ist in der Literatur bislang uneinheitlich beschrieben worden.³⁰ Im Allgemeinen steht der Begriff für Faktoren außerhalb einer Intervention, die die Wirkungen der Intervention beeinflussen können (und mit dieser interagieren können). Verwandte Konzepte sind **externe Validität (external validity)**, **Relevanz (relevance)**, **Anwendbarkeit (applicability)** oder **Indirektheit (indirectness)**.³² Eine auf systematische Reviews bezogene Definition der Übertragbarkeit (transferability) ist das Ausmaß, in dem sich der Kontext der Fragestellung des Reviews und der Kontext der Studien, deren Daten zum Ergebnis des Reviews beigetragen haben, in Bezug auf a priori definierte Merkmale unterscheiden.³⁰

Die bislang verfügbaren Instrumente zur Begutachtung der Übertragbarkeit der Ergebnisse wissenschaftlicher Arbeiten wurden größtenteils zunächst für die Begutachtung einzelner Primärstudien entwickelt,^{33,34} d. h. nicht für systematische Reviews. Mit dem **TRANSFER Approach** ist ein umfassender Ansatz für die Begutachtung der Übertragbarkeit (transferability) der Ergebnisse von systematischen Reviews vorgestellt worden, der sich jedoch noch in der Entwicklung befindet.³⁰

2.5 Metaanalysen

Eine **Metaanalyse** ist die quantitative Zusammenfassung der Ergebnisse von zwei oder mehr Studien mit statistischen Methoden.² Die Planung und Durchführung von Metaanalysen, aber auch ihre Bewertung, erfordert eine angemessene methodische Expertise. Metaanalysen können sehr komplex sein, und ihre Methoden werden stetig weiterentwickelt. So gibt es neben den klassischen **paarweisen Metaanalysen (pairwise metaanalyses)**, in denen die Ergebnisse von zwei Interventionen bzw. Maßnahmen verglichen werden, inzwischen **Netzwerkmetaanalysen (network metaanalyses, NMA)**, in denen die Ergebnisse von mehr als zwei Interventionen bzw. Maßnahmen in einer Analyse verglichen werden, und **Komponentennetzwerkmetaanalysen (component network metaanalyses, CNMA)**, in denen die einzelnen Komponenten von mehrere Komponenten umfassenden Interventionen in einer Netzwerkmetaanalyse verglichen werden können. Zu unterscheiden sind zudem **Metaanalysen mit aggregierten Studiendaten von**

Metaanalysen mit individuellen Teilnehmendendaten (individual participant data (IPD) metaanalyses).

Die Durchführung von Metaanalysen unterliegt, wie alle Review-Methoden, bestimmten Anforderungen. Nicht immer ist es sinnvoll, die Ergebnisse von Studien in einer Metaanalyse zusammenfassen. Dies ist insbesondere dann der Fall, wenn die Studien, deren Ergebnisse zusammengefasst werden sollen, **klinisch heterogen** sind, d.h., sich in ihren Populationen, Interventionen, Vergleichen und/oder Endpunkten relevant unterscheiden.

Die geplanten **Metaanalyse-Methoden** sollten a priori, d.h. im Zuge der Erstellung des Review-Protokolls, festgelegt werden. Zu berücksichtigen sind hierbei u. a. die Wahl des **Metaanalyse-Modells** („fixed effect“ oder „random effects“) und die Methoden zur Überprüfung der Metaanalyse-Ergebnisse auf **statistische Heterogenität**. Mit der Durchführung von **Heterogenitäts-Tests** kann ermittelt werden, ob die Unterschiede zwischen den Ergebnissen der in eine Metaanalyse eingeschlossenen Studien größer sind als zufällig zu erwarten wäre. Die Durchführung und Interpretation einer Metaanalyse mit heterogenen Studien kann problematisch sein. Bei vorliegender Heterogenität sollten deshalb zusätzliche Untersuchungen, z. B. **Sensitivitätsanalysen**, durchgeführt werden, um, sofern möglich, Erklärungen für die Heterogenität zu finden. Bestehen Zweifel, ob sich Studienergebnisse sinnvoll statistisch zusammenfassen lassen oder kann eine ausgeprägte statistische Heterogenität durch Sensitivitätsanalysen nicht erklärt werden, sollte auf die Metaanalyse verzichtet werden und erläutert werden, warum keine Metaanalyse erfolgte.

Die unzureichende Planung und Durchführung, aber auch die unzureichende Bewertung einer Metaanalyse kann zu falschen Ergebnissen bzw. zu Fehleinschätzungen ihrer Ergebnisse und damit zu falschen Schlüssen z. B. zur Wirksamkeit und Sicherheit einer Intervention führen. Empfohlen wird daher, wann immer möglich, eine in Metaanalyse-Methoden erfahrene Statistiker*in/einen in Metaanalyse-Methoden erfahrenen Statistiker in die Planung, Durchführung und Bewertung eines systematischen Reviews einzubeziehen. **Weiterführende Informationen** zu wesentlichen methodischen Aspekten von Metaanalysen finden sich im **Anhang**.

2.6 Vertrauenswürdigkeit der Evidenz

Neben der methodischen Qualität bzw. dem Biasrisiko sind für das Vertrauen in die Ergebnisse eines systematischen Reviews weitere Aspekte von Bedeutung. Diese können mit **GRADE (Grading of Recommendations Assessment, Development and Evaluation)**³⁵ strukturiert bewertet werden, einem Ansatz für eine standardisierte, transparente Einschätzung der Vertrauenswürdigkeit der (in einem systematischen Review zusammengefassten) Evidenz. Die GRADE-Bewertung wird nicht für ein Ergebnis einer einzelnen Studie, sondern für ein aus der Zusammenfassung der Ergebnisse mehrerer Studien generiertes Ergebnis durchgeführt. Zudem bezieht sie sich immer auf ein konkretes Ergebnis, wie z.B. das Ergebnis einer Metaanalyse zu einem spezifischen Endpunkt. In die GRADE-Bewertung fließen neben dem Biasrisiko unter anderem eine **unzureichende Präzision** (ein weites Konfidenzintervall), **Inkonsistenz** (Heterogenität, nicht übereinstimmende Ergebnisse), **Indirektheit** (Diskrepanzen zwischen der Fragestellung und der verfügbaren Evidenz in Bezug auf die Population, Intervention, Vergleichsgruppe und den Endpunkt) und **Disseminationsbias** ein. Die GRADE-Bewertung ist daher kein alternativer Ansatz zur Bewertung des Biasrisikos, sondern vielmehr ein umfassender Ansatz zur Bewertung der Vertrauenswürdigkeit von Evidenz in einem Review, der das Biasrisiko als einen von mehreren Aspekten beinhaltet. Weiterführende Informationen und Ressourcen zu GRADE sind u. a. auf den Webseiten des **Freiburg GRADE Center**³⁶ verfügbar.

3 DIE KRITISCHE BEWERTUNG VON SYSTEMATISCHEN REVIEWS

3.1 Einführung und Übersicht

Systematische Reviews können in ihrer Qualität variieren und ihre Ergebnisse und Schlussfolgerungen können mit einem Biasrisiko behaftet sein.³⁷ Wie zuvor beschrieben (s. Kapitel 2), sollte den Ergebnissen und Schlussfolgerungen eines systematischen Reviews nur dann vertraut werden, wenn er von guter Qualität ist und ein geringes Biasrisiko aufweist. Eine Vielzahl empirischer Untersuchungen hat gezeigt, dass viele systematische Reviews zu Fragestellungen aus allen Bereichen der Gesundheitsversorgung methodische Mängel und/oder ein Biasrisiko aufweisen.³⁸⁻⁴⁰ Entsprechend wichtig ist es, systematische Reviews vor der Nutzung ihrer Ergebnisse grundsätzlich kritisch zu bewerten.

Unter der kritischen Bewertung eines systematischen Reviews wird in diesem Manual die kriteriengestützte **Bewertung seiner methodischen Qualität** und/oder des **Biasrisikos seiner Ergebnisse** verstanden. **Der Fokus liegt dabei auf der Bewertung der methodischen Qualität und des Biasrisikos in systematischen Reviews zur Wirksamkeit und Sicherheit von therapeutischen Interventionen.** Weiterführende Informationen zu den PRISMA 2020-Ressourcen als Leitfaden für die Berichterstattung von systematischen Reviews zu therapeutischen Interventionen finden sich in **Kapitel 6**.

Für die kritische Bewertung von systematischen Reviews steht eine Vielzahl von **Bewertungsinstrumenten (critical appraisal tools)** zur Verfügung. Der überwiegende Teil der Instrumente beschreibt in Form einer Checkliste, welche Aspekte eines systematischen Reviews begutachtet werden sollten. Viele der verfügbaren Checklisten sind jedoch nicht oder nicht ausreichend validiert. Manche sind sehr umfangreich und damit wenig praktikabel, und manche bewerten neben der methodischen Qualität bzw. dem Biasrisiko auch die Berichtqualität.

Tabelle 1 zeigt eine **Übersicht über eine Auswahl verschiedener Bewertungsinstrumente für systematische Reviews**, die im Rahmen der Erstellung der initialen Version des Manuals auf den Webseiten führender Institutionen im Bereich der Evidenzbasierten Medizin oder in den Referenzlisten bereits bekannter Bewertungsinstrumente^{18,41-44} identifiziert wurden.

Tab. 1: Bewertungsinstrumente für systematische Reviews (Auswahl)

Bezeichnung	Veröffentlicht (Jahr)	Land	Link zum Instrument
A Measurement Tool to Assess systematic Reviews (AMSTAR) 2 ⁴⁵	2017	Kanada	https://amstar.ca/docs/AMSTAR-2.pdf
Risk of Bias in Systematic Reviews (ROBIS) ⁴⁶	2016	UK	https://www.bmj.com/content/358/bmj.j4008.full.pdf
Health Evidence Quality Assessment Tool	2016	Kanada	https://www.healthevidence.org/documents/our-appraisal-tools/QA_Tool&Dictionary_10Nov16.pdf
Documentation Appraisal Review Tool (DART)	2015	USA	http://www.wjgnet.com/2308-3840/coretip/v3/i3/142.htm
An Approach to Assess Confidence in Findings from Qualitative Evidence Syntheses (GRADE-CERQual) ⁴⁷	2015	International	https://implementationscience.biomedcentral.com/articles/10.1186/s13012-017-0688-3
Quality Assessment of Systematic Reviews and Meta-Analyses	2014	US	https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/sr_ma
Critical Appraisal of a Metaanalysis or Systematic Review	2014	Niederlande	https://www.cebma.org/wp-content/uploads/Critical-Appraisal-Questions-for-a-SR-or-MA.pdf
Specialist Unit for Review Evidence (SURE)	2013	UK	https://www.cardiff.ac.uk/specialist-unit-for-review-evidence/resources/critical-appraisal-checklists
Critical Appraisal Skills Programme (CASP)	2013	UK	https://casp-uk.net/casp-tools-checklists/
Quality Criteria Checklist: Review Article	2012	USA	http://www.adaevidencelibrary.com/files/Docs/2012_Jan_EA_Manual.pdf
Revised-AMSTAR (R-AMSTAR). ⁴⁸	2010	USA	http://www.perosh.eu/wp-content/uploads/2014/02/R-AMSTAR-Checklist-OSH-Evidence.pdf
A Measurement Tool to Assess systematic Reviews (AMSTAR) ^{18,41-44}	2007	Kanada	https://amstar.ca/index.php (s. AMSTAR 2)
Overview Quality Assessment Questionnaire (OQAQ) ^{49,50}	1991	Kanada	https://www.researchgate.net/figure/Overview-Quality-Assessment-Questionnaire_tbl1_51874218
Critical Appraisal Sheet	k.A.	UK	http://www.cebm.net/critical-appraisal/
Health Evidence Bulletins - Wales: Critical Appraisal of a systematic review (HEBW)	k.A.	UK	http://hebw.cf.ac.uk/projectmethod/appendix5.htm
Critical Appraisal Checklist for Systematic Reviews and Research Syntheses	k.A.	Australien	https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Systematic_Reviews2017_0.pdf
Assessing the Credibility of the Systematic Review Process	k.A.	UK	http://guides.mclibrary.duke.edu/ebm/appraise
Critical Appraisal Checklist for a systematic review	k.A.	UK	http://www.gla.ac.uk/media/media_64047_en.pdf
Systematic review (of therapy) worksheet	k.A.	Kanada	https://www.colleaga.org/tools/systematic-review-therapy-critical-appraisal-worksheet
Methodology Checklist: Systematic Reviews and Meta-analyses - Scottish Intercollegiate Guidelines Network (SIGN)	k.A.	UK	https://www.sign.ac.uk/what-we-do/methodology/checklists/
Reviews and Meta-Analyses Checklist	k.A.	UK	http://bestbets.org/ca/pdf/review.pdf

k.A.: keine Angabe.

3.2 Die Bewertungsinstrumente AMSTAR 2 und ROBIS

In diesem Manual werden zwei Instrumente zur Bewertung von systematischen Reviews zu Interventionsstudien detaillierter vorgestellt: das **AMSTAR 2 Tool**⁵¹ und das **ROBIS Tool**⁴⁶. Beide Instrumente sind in der aktuellen Version (6.3) des Cochrane Handbook for Systematic Reviews of Interventions²⁵ als geeignete Bewertungsinstrumente für systematische Reviews genannt. Mit der Begründung eines Mangels an empirischer Evidenz wird dabei keines der Instrumente als „das bessere“ empfohlen. Das initiale AMSTAR Tool wird im Cochrane Handbook als drittes geeignetes Instrument genannt, jedoch verweisen die Autor*innen darauf, dass AMSTAR 2 als aktualisierte Version von AMSTAR entwickelt wurde, was dafür spricht, zukünftig bevorzugt AMSTAR 2 anstelle von AMSTAR zu verwenden. Eine pragmatische Sichtung der 13 seit 2019 publizierten **Cochrane Overviews** ergab, dass in 5/13 (38%) AMSTAR 2, in 4/13 (31%) ROBIS, in 3/13 (23%) AMSTAR und in einem Overview ein anderes Instrument verwendet wurde (Recherche in Cochrane Library, Juni 2022).

Folgend sind einige wesentliche **Gemeinsamkeiten und Unterschiede von AMSTAR 2 und ROBIS** zusammengefasst, die eine Hilfestellung für die Entscheidung für das eine oder andere Instrument bieten können. Zu berücksichtigen ist dabei, dass die Anzahl der verfügbaren empirischen Untersuchungen zu den beiden Instrumenten sehr begrenzt ist und sich die dargestellten Ergebnisse zumeist auf einzelne Untersuchungen beziehen. In den **Kapiteln 4 und 5** werden die beiden Instrumente einzeln und detaillierter vorgestellt.

AMSTAR 2 und ROBIS unterscheiden sich in ihrem **theoretischen Ansatz**. AMSTAR 2 wurde als Instrument für die Bewertung der methodischen Qualität entwickelt⁵¹, ROBIS als Instrument zur Bewertung des Biasrisikos⁴⁶ (s. Abschnitte 2.1 und 2.2 zu den Begriffen methodische Qualität und Biasrisiko). Dies zeigt sich insbesondere in einem direkteren und expliziteren Zuschnitt von ROBIS auf die Einschätzung der Auswirkungen nicht oder unzureichend erfüllter methodischer Aspekte auf das Risiko für Bias.

Die beiden Instrumente unterscheiden sich zudem darin, welche **Arten von systematischen Reviews bzw. Studientypen** mit ihnen bewertet werden können. AMSTAR 2 wurde ausschließlich für die Bewertung von systematischen Reviews zu (therapeutischen) Interventionsstudien konzipiert, mit ROBIS können neben systematischen Reviews zu Interventionsstudien auch

systematische Reviews zu anderen Studientypen (d. h. insbesondere zu ätiologischen, diagnostischen und prognostischen Fragestellungen) bewertet werden.

AMSTAR 2 und ROBIS weisen Gemeinsamkeiten in den in ihnen adressierten **methodischen Konstrukten** auf. So ergab eine Untersuchung von Swierz et al.⁵², dass 70% der Items von AMSTAR 2 und ROBIS dieselben oder ähnliche methodische Konstrukte adressieren. Beide Instrumente adressieren jedoch auch spezifische Aspekte, die in dem jeweils anderen Instrument nicht enthalten sind.

In Untersuchungen zur **Validität und Reliabilität** der beiden Instrumente wurden beide Instrumente als (für ihre Verwendung ausreichend und vergleichbar) valide bewertet.^{53,54} Die Intertester-Reliabilität wurde für ROBIS als „fair“⁵³ („ausreichend“), für AMSTAR 2 als „moderat“⁵⁴ und insgesamt als vergleichbar bewertet.^{55,56}

Die **Komplexität und der Anspruch an die Bewertung** mit ROBIS wird als höher betrachtet,^{55,57} d. h. sie erfordert eine höhere methodische Expertise im Bereich von systematischen Reviews. Die Bewertung mit AMSTAR 2 wird als vergleichsweise einfacher und nutzerfreundlicher empfunden.⁵⁵ Entsprechend ist die Bewertung mit ROBIS ggf. in erster Linie erfahrenen Wissenschaftler*innen und Methodiker*innen zu empfehlen.^{46,55,57} Auch der **zeitliche Aufwand der Bewertung** mit ROBIS wird als höher betrachtet.^{55,57}

4 DIE BEWERTUNG DER METHODISCHEN QUALITÄT VON SYSTEMATISCHEN REVIEWS MIT AMSTAR 2

4.1 Hintergrund

Das bislang wahrscheinlich am häufigsten verwendete Instrument zur Bewertung der methodischen Qualität von systematischen Reviews zu Interventionsstudien ist **AMSTAR (A MeaSurement Tool to Assess systematic Reviews)**.²⁵ Das von einer kanadischen Arbeitsgruppe unter Federführung von Beverley Shea von der University of Ottawa, Kanada, entwickelte Instrument wurde initial 2007 publiziert⁴². AMSTAR wurde auf Grundlage des 1991 von Oxman und Guyatt vorgestellten OQAQ (Overview Quality Assessment Questionnaire) Instrument^{49,50} und der Checkliste von Sacks et al.⁵⁸ entwickelt. Inzwischen wurde das Instrument umfassend überarbeitet. Die neue, derzeit aktuelle Version wurde 2017 unter dem Namen **AMSTAR 2** publiziert⁵¹.

AMSTAR 2 kann sowohl online als auch offline ausgefüllt und verwendet werden. Das Instrument sowie umfassende Informationen und Ressourcen zur Entwicklung und Anwendung sind auf einer eigenen **Website** verfügbar. Zurzeit sind keine validierten deutschen Übersetzungen des Instruments und der Ressourcen verfügbar.

Die in diesem Manual dargestellten Informationen zu **AMSTAR 2** basieren im Wesentlichen (und wo nicht anders ausgewiesen) auf den folgenden zu dem Tool verfügbaren Ressourcen, die als vorbereitende Lektüre vor der Anwendung des Tools dringend empfohlen sind. Die vollständigen bibliographischen Angaben finden sich in der Literaturliste am Ende des Manuals:

- Shea et al. 2017: **AMSTAR 2 Tool**⁵⁹
- Shea et al. 2017: **AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both**⁵¹
- Shea et al. 2017: **AMSTAR 2 Guidance Document**⁶⁰

4.2 Welche Arten von systematischen Reviews können mit AMSTAR 2 bewertet werden?

Während AMSTAR ursprünglich nur für die kritische Bewertung von systematischen Reviews zu **randomisierten kontrollierten Interventionsstudien (RCTs)** entwickelt wurde, können mit AMSTAR 2 systematische Reviews zu RCTs und/oder **nicht-randomisierten kontrollierten**

Interventionsstudien (NRSI) bewertet werden. Einige Fragen des Instruments beziehen sich gleichermaßen auf beide Studientypen und werden entsprechend unabhängig vom Studientyp bewertet, für andere gibt es getrennte Bewertungsabschnitte, von denen jeweils einer oder beide zu bewerten sind, je nachdem, ob der zu bewertende Review RCTs oder NRSI oder RCTs und NRSI enthält.

4.3 Wie ist AMSTAR 2 aufgebaut?

AMSTAR 2 ist eine Checkliste mit 16 Items, die als Fragen formuliert sind. Tabelle 2 zeigt eine Übersicht über die 16 Items. Zu jedem Item wurde eine Zusammenstellung von Kriterien formuliert, die für eine positive Bewertung („für ja“) erfüllt sein sollten und die Bewertung erleichtern sollen. Neben Kriterien, die für eine „ja“-Bewertung erfüllt sein sollten, gibt es bei einigen Items eine Unterteilung der Kriterien in „für teilweise ja“ und „für ja“, in einem Fall (Item 1) ein „optional (empfohlen)“.

Tab. 2: Die 16 Items/Fragen der AMSTAR 2 Checkliste (eigene Übersetzung)

Nr	Item/ Frage
1	Beinhalten die Forschungsfragen und Einschlusskriterien für den Review die PICO-Komponenten?
2	Ist in dem Review-Bericht explizit angegeben, dass die Review-Methoden vor der Durchführung des Reviews festgelegt wurden, und werden in dem Bericht jegliche Abweichungen vom Protokoll begründet?
3	Erläutern die Review-Autor*innen ihre Auswahl der Studientypen für den Einschluss in den Review?
4	Haben die Review-Autor*innen für ihre Literatursuche eine umfassende Suchstrategie angewandt?
5	Haben die Review-Autor*innen die Studienselektion doppelt durchgeführt?
6	Haben die Review-Autor*innen die Datenextraktion doppelt durchgeführt?
7	Haben die Review-Autor*innen eine Liste der ausgeschlossenen Studien bereitgestellt und die Ausschlüsse begründet?
8	Haben die Review-Autor*innen die eingeschlossenen Studien angemessen detailliert beschrieben?
9	Haben die Review-Autor*innen eine angemessene Technik für die Bewertung des Risikos für Bias (RoB) in den einzelnen, in den Review eingeschlossenen Studien angewandt?
10	Haben die Review-Autor*innen die Finanzierungsquellen der in den Review eingeschlossenen Studien angegeben?
11	Sofern eine Meta-Analyse durchgeführt wurde: Haben die Review-Autor*innen angemessene Methoden für die statistische Zusammenfassung der Ergebnisse angewandt?
12	Sofern eine Meta-Analyse durchgeführt wurde: Haben die Review-Autor*innen den potenziellen Einfluss des RoB der einzelnen Studien auf die Ergebnisse der Meta-Analyse bzw. andere Formen der Evidenzsynthese bewertet?
13	Haben die Review-Autor*innen dem RoB der einzelnen Studien im Rahmen der Interpretation/Diskussion der Ergebnisse des Reviews Rechnung getragen?
14	Haben die Review-Autor*innen für jegliche in den Ergebnissen des Reviews beobachtete Heterogenität zufriedenstellende Erklärungen bereitgestellt und die Heterogenität diskutiert?
15	Sofern eine quantitative Synthese durchgeführt wurde: Haben die Review-Autor*innen eine angemessene Untersuchung von Publikations-Bias („Small Study Bias“) durchgeführt und ihren wahrscheinlichen Einfluss auf die Ergebnisse des Reviews diskutiert?
16	Haben die Review-Autor*innen jegliche potenzielle Quellen für Interessenkonflikte, einschließlich jeglicher Finanzierung, die sie für die Durchführung des Reviews erhalten haben, angegeben?

Die **Bewertungsoptionen** für die AMSTAR 2 Items sind die folgenden. Für jedes Item gibt es mindestens zwei Antwortoptionen („ja“, „nein“).

- **„ja“** („yes“)
- **„teilweise ja“** („partial yes“)
- **„teilweise nein“** („partial no“)
- **„nein“** („no“)
- **„enthält nur RCTs“** („includes only RCTs“)
- **„enthält nur NRSI“** („includes only NRSI“)
- **„keine Meta-Analysen durchgeführt“** („no meta-analysis conducted“)

Abbildung 1 zeigt das AMSTAR 2 Item 9 als Beispiel für die Gestaltung der AMSTAR-Items, der Kriterien für die Bewertung, der Bewertungsoptionen sowie der Möglichkeit der getrennten Bewertung für Reviews mit RCTs und/oder NRSI.

AMSTAR 2 Item 9: Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?		
RCTs		
For Partial Yes, must have assessed RoB from:	For Yes, must also have assessed RoB from:	
<input type="checkbox"/> unconcealed allocation, and	<input type="checkbox"/> allocation sequence that was not truly random, and	<input type="checkbox"/> Yes
<input type="checkbox"/> lack of blinding of patients and assessors when assessing outcomes (unnecessary for objective outcomes such as all-cause mortality)	<input type="checkbox"/> selection of the reported result from among multiple measurements or analyses of a specified outcome	<input type="checkbox"/> Partial yes
		<input type="checkbox"/> No
		<input type="checkbox"/> Includes only NRSI
NRSI		
For Partial Yes, must have assessed RoB:	For Yes, must also have assessed RoB:	
<input type="checkbox"/> from confounding, and	<input type="checkbox"/> methods used to ascertain exposures and outcomes, and	<input type="checkbox"/> Yes
<input type="checkbox"/> from selection bias	<input type="checkbox"/> selection of the reported result from among multiple measurements or analyses of a specified outcome	<input type="checkbox"/> Partial Yes
		<input type="checkbox"/> No
		<input type="checkbox"/> Includes only RCTs

Abb. 1: AMSTAR 2 - Item/Frage 9

4.4 Wie wird die AMSTAR 2-Bewertung durchgeführt?

Alle Items sind zunächst einzeln zu bewerten. Hierzu wird überprüft, welche der zu jedem Item formulierten Kriterien erfüllt sind; hieraus kann anschließend die Bewertung für das jeweilige Item abgeleitet werden. Alle Bewertungen lassen sich durch das Setzen eines Häkchens (oder Kreuzes) in die hierfür vorgesehenen Kästchen dokumentieren.

Aus den Bewertungen der AMSTAR 2 Items kann eine **Bewertung des Gesamt-Vertrauens in die Ergebnisse** des systematischen Reviews abgeleitet werden. Hierzu hat die AMSTAR-Autor*innengruppe sieben der 16 Items (Fragen 2, 4, 7, 9, 11, 13, 15) als „**kritische Items**“ („**critical items**“) vorgeschlagen. Alle übrigen Fragen werden als „**nicht-kritische Items**“ („**non-critical items**“) bezeichnet. Je nachdem, wie viele kritische und nicht kritische Fragen mit „Nein“ (bzw. nicht mit „Ja“ oder „teilweise Ja“) bewertet wurden, kann das Gesamt-Vertrauen in die Ergebnisse

des Reviews als „hoch“, „moderat“, „niedrig“ oder „kritisch niedrig“ bewertet werden. Die AMSTAR-Gruppe weist in Ihrem **AMSTAR 2 Article**⁵¹ ausdrücklich darauf hin, dass die Auswahl der „kritischen Items“ als Vorschlag zu betrachten ist und Bewerter*innen-Teams diese Auswahl ggf. auch verändern können, je nachdem, welche Items sie für den von ihnen zu bewertenden Review für am wichtigsten erachten. Abb. 2 zeigt die AMSTAR 2-Kriterien für das Urteil über das Gesamt-Vertrauen in die Ergebnisse eines Reviews.

Von der Verwendung eines Gesamt-Scores aus den einzelnen Bewertungen der Items wird explizit abgeraten.⁵¹ Vielmehr sollen Anwender*innen des Tools den potenziellen Einfluss eines „negativen“ Bewertungsergebnisses für jedes einzelne Item bewerten.

<p>Hoch</p> <ul style="list-style-type: none"> • <i>Keine oder eine kritische Schwäche:</i> Der systematische Review bietet eine verlässliche und umfassende Zusammenfassung der Ergebnisse der verfügbaren Studien zu der interessierenden Frage. <p>Moderat</p> <ul style="list-style-type: none"> • <i>Mehr als eine nicht kritische Schwäche*:</i> Der systematische Review hat mehr als eine Schwäche, aber keine kritischen Schwächen. Möglicherweise bietet es eine akkurate Zusammenfassung der Ergebnisse der verfügbaren, in den Review eingeschlossenen Studien. <p>Niedrig</p> <ul style="list-style-type: none"> • <i>Ein kritischer Mangel mit oder ohne nicht kritische Schwächen:</i> Der Review hat eine kritische Schwäche und bietet möglicherweise keine akkurate und umfassende Zusammenfassung der verfügbaren Studien zu der interessierenden Frage. <p>Kritisch niedrig</p> <ul style="list-style-type: none"> • <i>Mehr als eine kritische Schwäche mit oder ohne nicht kritische Schwächen:</i> Der Review hat mehr als eine kritische Schwäche. Es sollte nicht darauf vertraut werden, dass er eine akkurate und umfassende Zusammenfassung der verfügbaren Studien bietet. <p>*Mehrere nicht kritische Schwächen können das Vertrauen in den Review mindern, und es kann angemessen sein, das Gesamturteil von moderat zu niedrigem Vertrauen herabzustufen.</p>

Abb. 2: AMSTAR 2 - Bewertung des Gesamt-Vertrauens in die Ergebnisse des Reviews^{51(p6)} (eigene Übersetzung)

5 DIE BEWERTUNG DES BIASRISIKOS IN SYSTEMATISCHEN REVIEWS MIT ROBIS

5.1 Hintergrund

Ein weiteres neueres Instrument für die kritische Bewertung von systematischen Reviews ist das **ROBIS (Risk of Bias in Systematic Reviews) Tool**. Es wurde 2016 von einer Gruppe internationaler Wissenschaftler*innen unter Federführung von Penny Whiting und weiteren Wissenschaftler*innen der University of Bristol, Großbritannien, publiziert⁴⁶. In Abgrenzung zu AMSTAR 2, das als Instrument zur Bewertung der methodischen Qualität von systematischen Reviews entwickelt wurde, wurde ROBIS explizit als Instrument zur Bewertung des Biasrisikos in systematischen Reviews entwickelt (s. 2.1 und 2.2 zur Erklärung der Begriffe „methodische Qualität“ und „Biasrisiko“).

Das ROBIS Tool ist als [pdf-Datei](#)⁶¹ frei verfügbar. Das Instrument sowie umfassende Informationen und Ressourcen zu seiner Entwicklung und Anwendung sind auf den [Websites](#)⁶² der Bristol Medical School (University of Bristol, UK) verfügbar. Zurzeit sind keine deutschen Übersetzungen des Instruments und der weiterführenden Ressourcen verfügbar.

Die in diesem Manual dargestellten Informationen zu **ROBIS** basieren im Wesentlichen (und wo nicht anders ausgewiesen) auf den folgenden zu dem Tool verfügbaren Ressourcen, die als vorbereitende Lektüre vor der Anwendung des Tools dringend empfohlen sind. Die vollständigen bibliographischen Angaben finden sich in der Literaturliste am Ende des Manuals:

- Whiting et al. 2015: [ROBIS Tool](#)⁶¹
- Whiting et al. 2016: [ROBIS: A new tool to assess risk of bias in systematic reviews was developed](#)⁴⁶
- Whiting et al. 2017: [ROBIS Guidance](#)⁶³

Weitere [Ressourcen zu ROBIS](#)⁶² umfassen ein [ROBIS Training Webinar](#), eine [ROBIS Access Database](#) und zwei Vorlagen für die [tabellarische](#) und [graphische Darstellung](#) von ROBIS-Bewertungsergebnissen (s. hierzu auch **Kapitel 5.5**).

5.2 Welche Arten von systematischen Reviews können mit ROBIS bewertet werden?

ROBIS wurde entwickelt, um das Biasrisiko in systematischen Reviews mit Fragestellungen aus den Bereichen Interventionen, Ätiologie, Diagnose und Prognose zu bewerten.

5.3 Wie ist ROBIS aufgebaut?

Die Bewertung mit ROBIS erfolgt in drei **Phasen** (Teile):

- **Phase 1: Einschätzung der Relevanz** (optional)
- **Phase 2: Identifizierung von Bedenken bezüglich des Review-Prozesses**
- **Phase 3: Beurteilung des Biasrisikos**

ROBIS ist in Teilen ein domänenbasiertes Tool. Phase 2, die Bewertung des Biasrisikos des Reviews, ist in vier Domänen unterteilt, die sich auf unterschiedliche Bias-Aspekte beziehen. Die Domänen sind:

- | |
|---|
| <p>Domäne 1: Studienauswahlkriterien</p> <p>Domäne 2: Identifizierung und Auswahl der Studien</p> <p>Domäne 3: Datenextraktion und kritische Bewertung der Studien</p> <p>Domäne 4: Synthese und Ergebnisse</p> |
|---|

Abb. 3: Die vier ROBIS-Domänen (Phase 2; eigene Übersetzung)

Zu jeder der vier Domänen wurden mehrere **Signalfragen (signalling questions)** formuliert, die die Bewertung erleichtern sollen. Die Bewertungsoptionen für die Signalfragen sind:

- „**ja**“ („yes“ → „Y“)
- „**wahrscheinlich ja**“ („probably yes“ → „PY“)
- „**wahrscheinlich nein**“ („probably no“ → „PN“)
- „**nein**“ („no“ → „N“)
- „**keine Informationen**“ („no information“ → „NI“)

Die Bewertungsoptionen für die Bewertung der Domänen und das Gesamturteil sind:

- **niedrig (low)**
- **hoch (high)**
- **unklar (unclear)**

Abbildung 4 zeigt exemplarisch die ROBIS-Domäne 2.

Domain 2: Identification and Selection of Studies		
Describe methods of study identification and selection (e.g. numbers of reviewers involved)		
2.1	Did the search include an appropriate range of databases/electronic sources for published and unpublished reports?	Y/PY/PN/N/NI
2.2	Were methods additional to database searching used to identify relevant reports?	Y/PY/PN/N/NI
2.3	Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	Y/PY/PN/N/NI
2.4	Were restrictions based on date, publication format, or language appropriate?	Y/PY/PN/N/NI
2.5	Were efforts made to minimise error in selection of studies?	Y/PY/PN/N/NI
Concerns regarding methods used to identify and/or select studies		LOW/HIGH/UNCLEAR
Rationale for concern:		

Abb. 4: ROBIS – Domäne 2⁶¹

5.4 Wie wird die ROBIS-Bewertung durchgeführt?

Die optionale **Phase 1 (Einschätzung der Relevanz)** der ROBIS-Bewertung ermöglicht die Bewertung der Relevanz der Frage des zu bewertenden Reviews in Bezug auf die **Zielfrage (target question)**, das heißt auf die Frage, auf die die Bewerter*innen eine Antwort suchen (z. B. für einen Overview of Reviews oder eine Leitlinienempfehlung). Der Abgleich der Zielfrage mit der **Reviewfrage** erfolgt bei systematischen Reviews zu Interventionsstudien durch Eingabe der entsprechenden Angaben in eine einfache, in die **PICO-Komponenten (Population, Intervention, Comparison, und Outcome)** unterteilte Tabelle, bei systematischen Reviews mit anderen Studientypen (bei ätiologischen, diagnostischen oder prognostischen Fragestellungen) in einer entsprechend angepassten Tabelle. Die zu beantwortende Frage lautet „Stimmt die Fragestellung des Reviews mit der Zielfrage überein?“. Die **Antwortoptionen** sind „**ja**“ („yes“), „**nein**“ („no“) oder „**unklar**“ („unclear“).

Die **Phase 2 (Identifizierung von Bedenken bezüglich des Review-Prozesses)** der ROBIS-Bewertung beinhaltet die Bewertung des Biasrisikos des Reviews. Hierzu werden zunächst die Signalfragen zu jeder der vier Domänen beantwortet und anschließend für jede Domäne eine

Bewertung des Biasrisikos als „niedrig“, „hoch“ oder „unklar“ dokumentiert. Ergänzend soll die Bewertung durch Hinzufügen einer Rationale („rationale for concern“) begründet werden.

Detaillierte Informationen, Hilfestellungen und Beispiele zur Beantwortung der Signalfragen und der Bewertung des Biasrisikos finden sich im **ROBIS Guidance Dokument**.⁶³

Weiterführende Informationen zur systematischen Literatursuche (Bewertung Domäne 2) finden sich im **„Manual Systematische Recherche für Evidenzsynthesen und Leitlinien“**.⁶⁴ Weiterführende Informationen zu den statistischen Verfahren von Metaanalysen und Disseminationsbias (Bewertung Domäne 4) finden sich im **Anhang**.

Die **Phase 3 (Beurteilung des Biasrisikos)** der ROBIS-Bewertung beinhaltet die **Zusammenführung der Bewertungsergebnisse** („niedrig“, „hoch“, „unklar“) **aus den vier Domänen der Phase 2** sowie die **Beantwortung von drei weiteren Signalfragen (A-C)**, die sich auf die Interpretation der Reviewergebnisse beziehen. Ihre Beantwortung führt zur **finalen Bewertung des Biasrisikos des Reviews als „niedrig“, „hoch“ oder „unklar“**. **Abbildung 5** zeigt die drei Signalfragen (A-C).

- A.** Wurden im Rahmen der Ergebnisinterpretation alle in den Domänen 1-4 (in Phase 2) identifizierten Bedenken bezüglich des Biasrisikos adressiert?
- B.** Wurde die Relevanz der identifizierten Studien für die Forschungsfrage des Reviews angemessen berücksichtigt?
- C:** Haben die Review-Autor*innen die Hervorhebung bestimmter Ergebnisse aufgrund ihrer statistischen Signifikanz vermieden?

Abb. 5: ROBIS – Phase 3, Signalfragen (eigene Übersetzung)

Die Antwortoptionen für die drei Signalfragen entsprechen denen für die Bewertungen der Signalfragen der Domänen in Phase 2 („ja“, „wahrscheinlich ja“, „wahrscheinlich nein“, „nein“, „keine Informationen“).

Basierend auf der Beantwortung der drei Signalfragen erfolgt die Bewertung des Gesamt-Biasrisikos des Reviews; die Bewertungsoptionen sind

- **niedriges Biasrisiko** (low risk of bias)
- **hohes Biasrisiko** (high risk of bias)
- **unklares Biasrisiko** (unclear risk of bias)

Folgend findet sich eine orientierende **Hilfestellung für die Gesamteinschätzung des Biasrisikos** des Reviews (**ROBIS Guidance Document**^{60,S.31,Tabelle18}; eigene Übersetzung). Im Guidance Document finden sich weitere, detailliertere Informationen.

Niedriges Biasrisiko	In Phase 2 der Bewertung wurden keine Bedenken bezüglich des Reviewprozesses identifiziert <i>oder</i> die identifizierten Bedenken wurden in den Schlussfolgerungen des Reviews angemessen berücksichtigt. Die Schlussfolgerungen werden durch die Evidenz gestützt und beinhalten die Berücksichtigung der Relevanz der eingeschlossenen Studien für die Fragestellung des Reviews.
Hohes Biasrisiko	In Phase 2 der Bewertung wurden keine Bedenken identifiziert, die in den Schlussfolgerungen des Reviews nicht berücksichtigt wurden <i>oder</i> die Schlussfolgerungen werden durch die Evidenz nicht gestützt <i>oder</i> die Schlussfolgerungen beinhalten nicht die Berücksichtigung der Relevanz der eingeschlossenen Studien für die Fragestellung des Reviews.
Unklares Biasrisiko	Die verfügbaren Informationen sind unzureichend für ein Urteil über das Biasrisiko.

Abb. 6: ROBIS - Gesamteinschätzung des Biasrisikos des Reviews⁶⁰ (eigene Übersetzung)

5.5 Wie wird die ROBIS-Bewertung dokumentiert und präsentiert?

Die Entwickler*innen von ROBIS empfehlen, die Bewertung der einzelnen Domänen (aus Phase 2) tabellarisch darzustellen. Weiterführende Angaben zur tabellarischen und einer möglichen graphischen Darstellung des Verzerrungspotenzials sind in den **ROBIS-Ressourcen**⁶² verfügbar: Neben der tabellarischen Darstellung sollte die (möglicherweise) bestehende Ergebnisverzerrung auch narrativ beschrieben und diskutiert werden.

6 STANDARDS FÜR DIE BERICHTERSTATTUNG ZU SYSTEMATISCHEN REVIEWS – PRISMA 2020

In **Kapitel 2** wurde der Begriff der Berichtqualität erklärt und ein Hinweis auf Standards für die Berichterstattung von systematischen Reviews gegeben. Der bekannteste Standard für die Berichterstattung zu systematischen Reviews ist das **PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) Statement**.^{29,65} Der primäre Fokus des PRISMA Statements liegt auf der Berichterstattung von systematischen Reviews zu den Effekten von Interventionen, jedoch kann es auch als Basis für systematische Reviews zu anderen Fragestellungen verwendet werden. Das PRISMA Statement wurde initial 2009 von Moher et al.⁶⁵ publiziert und umfasste zunächst drei Dokumente: einen Artikel⁶⁵ (das **PRISMA Statement**), eine Checkliste⁶⁶ (die **PRISMA Checklist**) mit Aspekten, die in Berichten zu systematischen Reviews enthalten sein sollten, und eine Vorlage für ein Flussdiagramm⁶⁷ (**PRISMA Flow Diagram**) zur Dokumentation des Such- und Auswahlprozesses. Inzwischen wurde das PRISMA Statement umfassend überarbeitet und ergänzt. Die aktuelle Version des PRISMA Statements wurde 2020 unter dem Namen „**PRISMA 2020 Statement**“ publiziert²⁹ (seither werden die initialen PRISMA-Dokumente unter „PRISMA 2009“ geführt). Zu den Ergänzungen zählt die **Erweiterung des PRISMA-Flussdiagramms auf vier Versionen**, je zwei für neue und für aktualisierte systematische Reviews. Zudem gibt es **ergänzende Checklisten (PRISMA Extensions)** zur PRISMA 2020 Checkliste für bestimmte Aspekte oder Arten von systematischen Reviews, u. a. für Protokolle, Abstracts, Netzwerk-Metaanalysen und die Literatursuche.

Abbildung 7 zeigt eine Übersicht über die wesentlichen PRISMA-Dokumente und -Ressourcen. Diese sind über die eigene **PRISMA-Website**⁶⁸ verfügbar. Die PRISMA-Ressourcen werden kontinuierlich weiter ergänzt. Alternativ finden sich die PRISMA-Dokumente auf den **Websites des Equator Network**⁶⁹. Zurzeit sind keine deutschen Übersetzungen der aktuellen PRISMA 2020-Dokumente verfügbar.

„PRISMA 2020 Statement“ (Artikel)	http://dx.doi.org/10.1136/bmj.n71
Erläuterungen zu PRISMA 2020 (Artikel; „Explanation & Elaboration“)	http://dx.doi.org/10.1136/bmj.n160
PRISMA 2020 Checkliste	http://www.prisma-statement.org/documents/PRISMA_2020_checklist.docx
PRISMA 2020 Flussdiagramme	http://www.prisma-statement.org/PRISMAStatement/FlowDiagram
PRISMA Extensions (Ergänzungen)	http://www.prisma-statement.org/Extensions/

Abb. 7: PRISMA 2020 Statement: Wesentliche Dokumente und Ressourcen

7 LITERATUR

1. Kunz, R.; Khan, K.S.; Kleijnen, J.; Antes G. *Systematische Übersichtsarbeiten Und Meta-Analysen*. 3. Auflage; Hogrefe. 2009.
2. Deeks JJ, Higgins JPT, Altman DG (Herausgeber). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Herausgeber*innen). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.3 (aktualisiert Februar 2022). Cochrane, 2022. www.training.cochrane.org/handbook. [Letzter Zugriff 29.11.2022].
3. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ*. 2004;328(7430):39-41. doi:10.1136/bmj.328.7430.39.
4. Noyes J., Booth A., Cargo M., Flemming K., Harding A., Harris J., Garthside R., Hannes K., Pantoja T, Thomas J. Chapter 21: Qualitative evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Herausgeber*innen). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.3 (aktualisiert Februar 2022). Cochrane, 2022. www.training.cochrane.org/handbook. [Letzter Zugriff 29.11.2022].
5. CASP Qualitative Study Checklist. https://casp-uk.b-cdn.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf. [Letzter Zugriff 29.11.2022].
6. Lewin S, Booth A, Glenton C, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. *Implement Sci*. 2018;13(S1):2. doi:10.1186/s13012-017-0688-3
7. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Herausgeber*innen). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.3 (aktualisiert Februar 2022). Cochrane, 2022. www.training.cochrane.org/handbook. [Letzter Zugriff 29.11.2022].
8. Cochrane Library. <https://www.cochranelibrary.com/>. [Letzter Zugriff 29.11.2022].
9. Estevam JDA, Franco ESB, Kriebel CF, Peccin MS. Methodological Quality Analysis of Systematic Review for the Treatment of Rotator Cuff Disease. *Rev Bras Ortop*. 2021;56(4):485-489. doi:10.1055/s-0040-1710334.
10. Goldkuhle M, Narayan VM, Weigl A, Dahm P, Skoetz N. A systematic assessment of Cochrane reviews and systematic reviews published in high-impact medical journals related to cancer. *BMJ Open*. 2018;8(3):e020869. doi:10.1136/bmjopen-2017-020869.
11. Tricco AC, Tetzlaff J, Pham B, Brehaut J, Moher D. Non-Cochrane vs. Cochrane reviews were twice as likely to have positive conclusion statements: cross-sectional study. *J Clin Epidemiol*. 2009;62(4):380-386.e1. doi:10.1016/j.jclinepi.2008.08.008.
12. Hamel C, Michaud A, Thuku M, et al. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. *J Clin Epidemiol*. 2021;129:74-85. doi:10.1016/j.jclinepi.2020.09.041.
13. Garritty C, Gartlehner G, Nussbaumer-Streit B, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. *J Clin Epidemiol*. 2021;130:13-22. doi:10.1016/j.jclinepi.2020.10.007.
14. Cochrane Rapid Reviews Methods Group. <https://methods.cochrane.org/rapidreviews/>. [Letzter Zugriff 29.11.2022].
15. Elliott JH, Synnot A, Turner T, et al. Living systematic review: 1. Introduction—the why, what, when, and how. *J Clin Epidemiol*. 2017;91:23-30. doi:10.1016/j.jclinepi.2017.08.010.
16. Brooker J & LEN. *Guidance for the Production and Publication of Cochrane Living Systematic Reviews: Cochrane Reviews in Living Mode.*; Version Dezember 2019. https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201912_LSR_Revised_Guidance.pdf. [Letzter Zugriff 29.11.2022].
17. Akl EA, Meerpohl JJ, Elliott J, et al. Living systematic reviews: 4. Living guideline recommendations. *J Clin Epidemiol*. 2017;91:47-53. doi:10.1016/j.jclinepi.2017.08.009.
18. Hartling L, Chisholm A, Thomson D, Dryden DM. A descriptive analysis of overviews of reviews published between 2000 and 2011. *PLoS One*. 2012;7(11):e49667. doi:10.1371/journal.pone.0049667.
19. Ballard M, Montgomery P. Risk of bias in overviews of reviews: a scoping review of methodological guidance

- and four-item checklist. *Res Synth Methods*. 2017. doi:10.1002/jrsm.1229.
20. Caird J, Sutcliffe K, Kwan I, Dickson K, Thomas J. Mediating policy-relevant evidence at speed: are systematic reviews of systematic reviews a useful approach? *Evid Policy A J Res Debate Pract*. 2015;11(1):81-97.
 21. Miake-Lye IM, Hempel S, Shanman R, Shekelle PG. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Syst Rev*. 2016;5:28. doi:10.1186/s13643-016-0204-x.
 22. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. 2018;18(1):143. doi:10.1186/s12874-018-0611-x.
 23. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) - Ständige Kommission Leitlinien. AWMF-Regelwerk "Leitlinien". 2. Auflage 2020. <https://www.awmf.org/leitlinien/awmf-regelwerk.html>. [Letzter Zugriff 29.11.2022].
 24. O'Rourke B, Oortwijn W, Schuller T. The new definition of health technology assessment: A milestone in international collaboration. *Int J Technol Assess Health Care*. 2020;36(3):187-190. doi:10.1017/S0266462320000215.
 25. Pollock M, Fernandes RM, Becker LA, Pieper D HL. Chapter 5: Overviews of Reviews. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Herausgeber*innen). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.3 (aktualisiert Februar 2022). Cochrane, 2022. www.training.cochrane.org/handbook. [Letzter Zugriff 29.11.2022].
 26. PROSPERO. International Prospective Register of Systematic Reviews. (Website). <http://www.crd.york.ac.uk/PROSPERO/>. [Letzter Zugriff 29.11.2022].
 27. Higgins, J.; Lasserson, T.; Chandler, J.; Tovey, D.; Thomas, J.; Flemyng, E.; Churchill R. *Methodological Expectations of Cochrane Intervention Reviews (MECIR). Standards for the Conduct and Reporting of New Cochrane Intervention Reviews, Reporting of Protocols and the Planning, Conduct and Reporting of Updates. Version Februar 2022*. <https://community.cochrane.org/mecir-manual>. [Letzter Zugriff 29.11.2022].
 28. Cumpston, M.; Lasserson, T.; JChandler, J.; Page MJ. Chapter 3: Reporting the review. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Herausgeber*innen). *Cochrane Handbook for Systematic Reviews of Interventions*, Version 6.3 (aktualisiert Februar 2022). Cochrane, 2022. www.training.cochrane.org/handbook. [Letzter Zugriff 29.11.2022].
 29. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. March 2021:n71. doi:10.1136/bmj.n71.
 30. Munthe-Kaas H, Nøkleby H, Lewin S, Glenton C. The TRANSFER Approach for assessing the transferability of systematic review findings. *BMC Med Res Methodol*. 2020;20(1):1-22. doi:10.1186/s12874-019-0834-5.
 31. Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply?" *Lancet*. 2005;365(9453):82-93. doi:10.1016/S0140-6736(04)17670-8.
 32. Murad MH, Katabi A, Benkhadra R, Montori VM. External validity, generalisability, applicability and directness: a brief primer. *BMJ Evidence-based Med*. 2018;23(1):17-19. doi:10.1136/ebmed-2017-110800.
 33. Dyrvig A-K, Kidholm K, Gerke O, Vondeling H. Checklists for external validity: a systematic review. *J Eval Clin Pract*. 2014;20(6):857-864. doi:10.1111/jep.12166.
 34. Jung A, Balzer J, Braun T, Luedtke K. Identification of tools used to assess the external validity of randomized controlled trials in reviews: a systematic review of measurement properties. *BMC Med Res Methodol*. 2022;22(1):1-23. doi:10.1186/s12874-022-01561-5.
 35. Langer G, Meerpohl JJ, Perleth M, Gartlehner G, Kaminski-Hartenthaler A, Schünemann H. GRADE-Leitlinien: 1. Einführung – GRADE-Evidenzprofile und Summary-of-Findings-Tabellen. *Z Evid Fortbild Qual Gesundheitswes*. 2012;106(5):357-368. doi:10.1016/j.zefq.2012.05.017.
 36. *Freiburg GRADE Center*. <https://de.gradeworkinggroup.org/>. [Letzter Zugriff 29.11.2022].
 37. Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q*. 2016;94(3):485-514. doi:10.1111/1468-0009.12210.
 38. Eshaghpour A, Li A, Javidan AP, Chen N, Yang S, Crowther MA. Evaluating the quality of systematic reviews and meta-analyses published on direct oral anticoagulants in the past 5 years. *BMJ Evidence-Based Med*. 2021;26(4):167-171. doi:10.1136/bmjebm-2019-111326.

39. Matthias K, Rissling O, Pieper D, et al. The methodological quality of systematic reviews on the treatment of adult major depression needs improvement according to AMSTAR 2: A cross-sectional study. *Heliyon*. 2020;6(9):e04776. doi:10.1016/j.heliyon.2020.e04776.
40. Sathish M, Eswar R. Systematic Reviews and Meta-Analysis in Spine Surgery—How Good Are They in Methodological Quality? A Systematic Review. *Glob Spine J*. 2021;11(3):378-399. doi:10.1177/2192568220906810.
41. Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One*. 2007;2(12):e1350. doi:10.1371/journal.pone.0001350.
42. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol*. 2007;7:10. doi:10.1186/1471-2288-7-10.
43. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol*. 2009;62(10):1013-1020. doi:10.1016/j.jclinepi.2008.10.009.
44. Pieper D, Buechter R, Jerinic P, Eikermann M. Overviews of reviews often have limited rigor: a systematic review. *J Clin Epidemiol*. 2012;65(12):1267-1273. doi:10.1016/j.jclinepi.2012.06.015.
45. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:1-9. doi:10.1136/bmj.j4008.
46. Whiting P, Savović J, Higgins JPT, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225-234. doi:10.1016/j.jclinepi.2015.06.005.
47. Lewin S, Glenton C, Munthe-Kaas H, et al. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). *PLoS Med*. 2015;12(10):e1001895. doi:10.1371/journal.pmed.1001895.
48. Kung J, Chiappelli F, Cajulis OO, et al. From Systematic Reviews to Clinical Recommendations for Evidence-Based Health Care: Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. *Open Dent J*. 2010;4:84-91. doi:10.2174/1874210601004020084.
49. Oxman AD, Guyatt GH, Singer J, et al. Agreement among reviewers of review articles. *J Clin Epidemiol*. 1991;44(1):91-98. doi:10.1016/0895-4356(91)90205-N.
50. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol*. 1991;44(11):1271-1278.
51. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:1-9. doi:10.1136/bmj.j4008.
52. Swierz MJ, Storman D, Zajac J, et al. Similarities, reliability and gaps in assessing the quality of conduct of systematic reviews using AMSTAR-2 and ROBIS: systematic survey of nutrition reviews. *BMC Med Res Methodol*. 2021;21(1):261. doi:10.1186/s12874-021-01457-w.
53. Bühn S, Mathes T, Prengel P, et al. The risk of bias in systematic reviews tool showed fair reliability and good construct validity. *J Clin Epidemiol*. 2017;91:121-128. doi:10.1016/j.jclinepi.2017.06.019.
54. Lorenz RC, Matthias K, Pieper D, et al. A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. *J Clin Epidemiol*. 2019;114:133-140. doi:10.1016/j.jclinepi.2019.05.028.
55. Perry R, Whitmarsh A, Leach V, Davies P. A comparison of two assessment tools used in overviews of systematic reviews: ROBIS versus AMSTAR-2. *Syst Rev*. 2021;10(1):273. doi:10.1186/s13643-021-01819-x.
56. Pieper D, Puljak L, González-Lorenzo M, Minozzi S. Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and nonrandomized studies. *J Clin Epidemiol*. 2019;108:26-33. doi:10.1016/j.jclinepi.2018.12.004.
57. Gates M, Gates A, Duarte G, et al. Quality and risk of bias appraisals of systematic reviews are inconsistent across reviewers and centers. *J Clin Epidemiol*. 2020;125:9-15. doi:10.1016/j.jclinepi.2020.04.026.
58. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med*. 1987;316(8):450-455. doi:10.1056/nejm198702193160806.
59. AMSTAR 2 - Tool (Shea et al. 2017). <https://amstar.ca/docs/AMSTAR-2.pdf>. [Letzter Zugriff 29.11.2022].
60. AMSTAR 2. Guidance Document (Shea et al. 2017). <https://amstar.ca/docs/AMSTAR%202-Guidance->

- [document.pdf](#). [Letzter Zugriff 29.11.2022].
61. ROBIS Tool (Whiting et al. 2016). <https://www.bristol.ac.uk/media-library/sites/social-community-medicine/robis/ROBIS%201.2%20Clean.pdf>. [Letzter Zugriff 29.11.2022].
 62. ROBIS Tool. Resources (Whiting et al. 2016). <https://www.bristol.ac.uk/population-health-sciences/projects/robis/resources/>. [Letzter Zugriff 29.11.2022].
 63. ROBIS. Guidance (Whiting et al. 2016). <https://www.bristol.ac.uk/media-library/sites/social-community-medicine/robis/robisguidancedocument.pdf>. [Letzter Zugriff 29.11.2022].
 64. Cochrane Deutschland Stiftung, Institut für Evidenz in der Medizin, Institut für Medizinische Biometrie und Statistik, Freiburg, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften – Institut für Medizinisches Wissensmanagement, Ärztliches Zentrum für Qualität in der Medizin. *Manual Systematische Recherche für Evidenzsynthesen und Leitlinien*. Version 2.1 (Dezember 2020). doi:10.6094/UNIFR/174468. <https://www.cochrane.de/de/literaturrecherche>; <https://www.awmf.org/leitlinien/awmf-regelwerk/ll-entwicklung.html>; <https://www.aezq.de/aezq/publikationen/azq-partner#literaturrecherche>; <https://freidok.uni-freiburg.de/data/174468>. [Letzter Zugriff 29.11.2022].
 65. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339(jul21 1):b2535-b2535. doi:10.1136/bmj.b2535.
 66. PRISMA 2009 checklist (Moher et al 2009). <https://prisma-statement.org/documents/PRISMA%202009%20checklist.pdf>. [Letzter Zugriff 29.11.2022].
 67. PRISMA 2009 Flow diagram (Moher et al. 2009). <https://prisma-statement.org/documents/PRISMA%202009%20flow%20diagram.pdf>. [Letzter Zugriff 29.11.2022].
 68. PRISMA Statement (Website). <https://www.prisma-statement.org/>. [Letzter Zugriff 29.11.2022].
 69. Equator Network (Website). <https://www.equator-network.org/>. [Letzter Zugriff 29.11.2022].

ANHANG: WEITERFÜHRENDE INFORMATIONEN ZU METAANALYSEN UND DISSEMINATIONSBIAS

Metaanalysen

Was ist ein Forest Plot?

- Häufigste Abbildungsart zur zusammenfassenden Darstellung der einzelnen Studienergebnisse einer Metaanalyse
- Für jede Studie wird der geschätzte Therapieeffekt mit zugehörigem Konfidenzintervall dargestellt
- Gibt für jede Studie Auskunft über die (1) Stärke des Therapieeffekts = Lage des Effektschätzers, (2) Unsicherheit = Breite des Konfidenzintervalls, (3) Informationsgehalt = Gewicht der Studie/Dicke des Markerpunktes.
- Gibt (meistens) auch einen über alle Studien gepoolten Schätzwert mit Konfidenzintervall an.
- Zusätzlich kann ein Prädiktionsintervall angegeben werden. Im Unterschied zum Konfidenzintervall, das das Ausmaß der Unsicherheit beschreibt, mit dem die Schätzung behaftet ist, stellt das Prädiktionsintervall einen Bereich dar, in dem eine zukünftige Studie zu erwarten ist. Seine Länge nimmt daher mit zunehmender Anzahl Studien in der Metaanalyse nicht ab, sondern stabilisiert sich.

Was ist ein gemeinsamer Schätzer?

Ein gewichteter Mittelwert der geschätzten Therapieeffekte über alle Einzelstudien hinweg.

Gibt es bei Metaanalysen unterschiedliche Methoden?

Die zwei einfachsten und bekanntesten Methoden der Metaanalyse sind das Fixed Effect Modell (zunehmend auch Common Effect Modell genannt) und das Random Effects Modell. Sie unterscheiden sich in der Art der Gewichtung der Einzelstudien.

Was ist der Unterschied zwischen dem Fixed Effect Modell und dem Random Effects Modell?

Annahmen für das Fixed Effect Modell:

- allen Studien liegt, bis auf Zufallsschwankungen, der gleiche Therapieeffekt zugrunde
- Unterschiede in den Therapieeffekten sind ausschließlich auf Zufallsstreuung zurückzuführen
- es gibt dafür mehrere Gewichtungsmethoden

Annahmen für das Random Effects Modell:

- Therapieeffekt zwischen den Einzelstudien variiert natürlicherweise; Variation nach fest vorgegebener Verteilung (meistens Normalverteilung)

Wieso nimmt man nicht immer das Random Effects Modell?

Im Random Effects Modell erhalten kleinere Studien ein größeres Gewicht als im Fixed Effekt Modell, somit kommt es zu einer Vereinheitlichung der Gewichtung, was das Random Effects Modell anfälliger für „small study effects“ macht.

Inwiefern unterscheiden sich die Ergebnisse zwischen Fixed Effect und Random Effects Modell?

Wenn wenig oder keine statistische Heterogenität vorliegt, unterscheiden sich die Ergebnisse zwischen den beiden Modellen wenig oder gar nicht. Das Fixed Effect Modell kann somit als Spezialfall des Random Effects Modell angesehen werden für den Fall, dass keine Heterogenität vorliegt. Wenn (klinische) Heterogenität vermutet wird, sollte das Random Effects Modell gewählt werden, das Fixed Effect Modell würde ein zu enges Konfidenzintervall liefern. Große Unterschiede zwischen Fixed Effect Modell und Random Effects Modell können darauf hinweisen, dass sich die kleinen Studien systematisch von den großen unterscheiden.

Die Wahl des Modells erfolgt bei der Erstellung des Protokolls, nicht auf der Basis der späteren Ergebnisse: „The choice between a fixed effect and a random effects meta-analysis should never be made on the basis of a statistical test of heterogeneity“.¹

Welche Arten von Heterogenität gibt es?

Zu den verschiedenen Arten von Heterogenität zählen²:

- Klinische Heterogenität: verursacht durch Unterschiede in den PICO-Charakteristika der Primärstudien, wie etwa unterschiedliche Patientenkollektive, unterschiedliche Varianten von Prüf- oder Vergleichsintervention (z. B. hinsichtlich der Dosierung oder der Dauer), oder unterschiedliche Nachbeobachtungszeiten;
- Statistische Heterogenität: auf der Endpunktskala quantifizierbar, statistische Heterogenität kann u.a. mit Cochrane's Q, I^2 oder τ^2 gemessen werden (siehe unten);
- Heterogenität verursacht durch unterschiedliche Studientypen („design-related“);
- Heterogenität verursacht durch Disseminationsbias (insbesondere Publikationsbias).

Wie testet man Ergebnisse auf statistische Heterogenität?

Mit Hilfe der statistischen Heterogenität kann man entscheiden, ob es überhaupt Sinn macht, die einzelnen Studienergebnisse in einem gemeinsamen Schätzer zusammenzufassen. Folgende Heterogenitätstest können u.a. dabei angewendet werden¹:

Cochran's Q:

- $Q=0$, falls der Therapieeffekt in allen Studien gleich ist
- Q untersucht nur, ob statistisch signifikante Heterogenität vorhanden ist oder nicht (Ja/Nein)
- Q macht keine Angabe über das Ausmaß an Heterogenität, dafür ist I^2 besser.

Higgins' und Thompson's I^2 :

- $I^2=0\%$, d.h. Variabilität zwischen den Studien ist ausschließlich durch Zufallsschwankungen erklärbar
- $I^2=100\%$, d.h. Variabilität zwischen den Studien ist ausschließlich durch Unterschiede zwischen den Studien erklärbar
- I^2 misst den Überlappungsgrad der Konfidenzintervalle

- I^2 wächst mit Präzision (ist also abhängig von der Studiengröße); bei sehr großen Studien, die gleichzeitig ein enges Konfidenzintervalle aufweisen, geht I^2 zwangsläufig gegen 100%; das bedeutet jedoch nicht unbedingt einen klinisch relevanten Unterschied.

Interstudienvarianz τ^2 :

- τ^2 beschreibt Variabilität zwischen den Studien, hängt aber nicht wie I^2 von der Studiengröße ab
- dafür gibt es in der Literatur mehrere Schätzer², der nach DerSimonian und Laird (1986) gilt als der am meisten verwendete; heute wird jedoch der REML-Schätzer empfohlen.³
- die Wurzel von τ^2 , also τ , ist in der gleichen Einheit wie der Endpunkt, somit lässt sich Heterogenität quantifizieren; es sollte jedoch vorher festgelegt werden, ab wann eine bedeutsame Heterogenität auf der Endpunktskala vorliegt.

Was macht man bei (sehr) heterogenen Ergebnissen?

Das Random Effects Modell kann Heterogenität beschreiben, aber nicht erklären. Zur Erklärung von Heterogenität sind zusätzliche Untersuchungen notwendig. Subgruppenanalysen und Metaregressionen können helfen, die statistische Heterogenität zu erklären. Solche Analysen sollten idealerweise vorab spezifiziert werden. Heterogenität kann ggf. auch klinisch erklärbar sein, bei nicht erklärbarer erheblicher Heterogenität sollte auf eine Metaanalyse verzichtet werden.

Subgruppenanalysen

Es werden Subgruppen von Studien nach vorab spezifizierten dichotomen oder nominalen klinischen Faktoren (wie z.B. Geschlecht, Altersgruppen, Komorbidität, sofern diese in den Primärstudien getrennt berichtet werden) gebildet.¹ Dies Subgruppen können eventuell Unterschiede in der Effektgröße erklären.

Metaregression

Eine Metaregression ist eine Verallgemeinerung der Subgruppenanalyse, bei der auch kontinuierliche Kovariablen und Effekte mehrerer Kovariablen jeglichen Typs simultan untersucht werden können. Eine Metaregression sollte nur durchgeführt werden, wenn mindestens 10 Studien in die Metaanalyse eingeschlossen wurden. Es wird allerdings davor gewarnt, Kovariablen, die auf Patientenebene definiert sind, in einer Subgruppenanalyse oder Metaregressionsanalyse zu untersuchen, da es dabei zu einem Bias durch Confounding kommen kann.¹

Sensitivitätsanalyse

Im Unterschied zu Subgruppenanalyse und Metaregression, bei denen es um klinische Heterogenität geht, ist das Ziel einer Sensitivitätsanalyse, den Einfluss anderer Faktoren auf das Ergebnis der Metaanalyse zu untersuchen. Beispiele für Sensitivitätsanalysen sind der Ausschluss einzelner Studien aus der Analyse (z.B. aufgrund eines hohen Verzerrungsrisikos), um einen Einfluss auf das Ergebnis (Effektgröße) zu untersuchen¹ oder die Verwendung eines Fixed Effect Modells, wenn die Hauptanalyse mit dem Random Effects Model durchgeführt wird.

Disseminationsbias

Was ist Disseminationsbias?

Disseminationsbias (dissemination bias) ist der Oberbegriff einer ganzen Reihe von Biastypen, die auf eine selektive Verbreitung von Studienergebnissen zurückgehen.⁴ Dazu zählen:

- publication bias (Publikationsbias; Studien mit negativen Ergebnissen werden seltener oder gar nicht publiziert)
- language bias (Sprachbias; Studien mit negativen Ergebnissen werden eher in nationalsprachlichen Zeitschriften publiziert)
- time-lag bias (Studien mit negativen Ergebnissen werden zeitverzögert publiziert)
- citation bias (Zitationsbias; Studien mit negativen Ergebnissen werden seltener zitiert)

Wie kann man Disseminationsbias untersuchen?

Mit Hilfe von graphischen und/oder statistischen Tests lässt sich das Vorhandensein von Disseminationsbias „annäherungsweise“ untersuchen.

Funnel Plot (graphischer Test)

Der Funnel Plot ist ein Scatterplot, bei dem ein Maß für die Studiengröße, zum Beispiel der Standardfehler in umgekehrter Richtung, gegen die Effektgröße (z. B. ein relatives Risiko) abgetragen wird. Werden kleine Studien mit geringen oder unerwünschten Effekten nicht publiziert, so erkennt man das an einem Funnelplot an der durch ihr Fehlen entstehenden Asymmetrie (in der rechten Abbildung fehlen die weißen Kreise).

Die Untersuchung von Disseminationsbias mit Hilfe eines Funnel Plots macht jedoch erst ab einer Anzahl von mindestens 10 eingeschlossenen Studien in die Metaanalyse Sinn: „As a rule of thumb, tests for funnel plot asymmetry should be used only when there are at least 10 studies included in the metaanalysis, because when there are fewer studies the power of the tests is too low to distinguish chance from real asymmetry.“^{1,5}

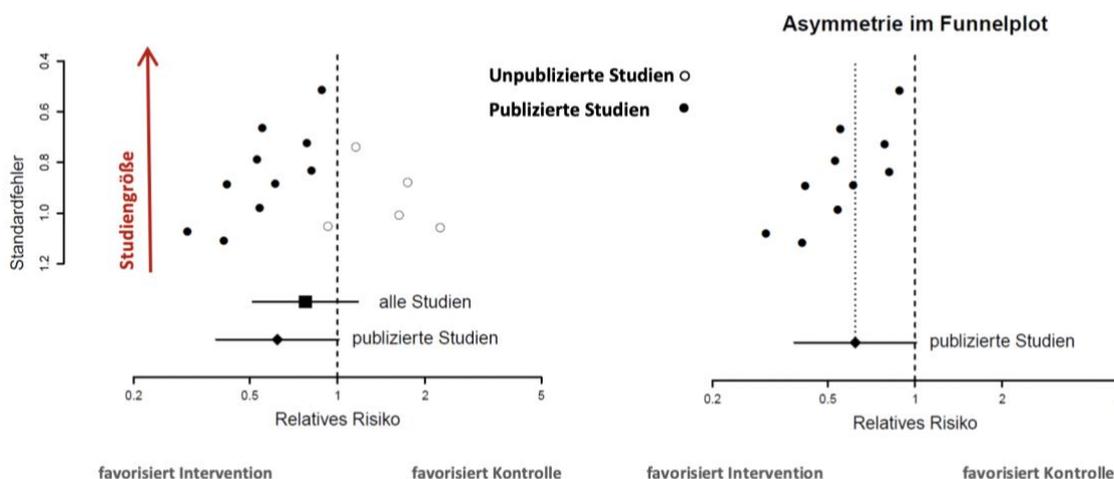


Abb. 8: Funnel Plot (Guido Schwarzer)

→ Effekt bei publizierten Studien im Vergleich zu allen Studien stärker ausgeprägt

Alternative Erklärungen für Asymmetrie im Funnel Plot:

- Klinische Heterogenität zwischen den Studien verursacht durch Unterschiede im Baseline-Risiko bei (stark) selektierten Patienten in kleinen Studien
- Sonstige Heterogenität: hohes Verzerrungsrisiko in kleinen Studien, was eine Überschätzung der Ergebnisse zur Folge haben kann („small study effects“)
- “Selective Outcome Reporting“: Statt des eigentlich geplanten primären Endpunkts wird ein anderer Endpunkt berichtet, falls sich bei diesem eine höhere Effektstärke zeigt.
- Zufall

Literatur

1. Deeks JJ, Higgins JPT, Altman DG (Herausgeber). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Herausgeber*innen). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.3 (aktualisiert Februar 2022). Cochrane, 2022. www.training.cochrane.org/handbook. [Letzter Zugriff 29.11.2022].
2. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I² in assessing heterogeneity may mislead. *BMC Med Res Methodol*. 2008;8(1):79. doi:10.1186/1471-2288-8-79.
3. Langan D, Higgins JPT, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res Synth Methods*. 2019;10(1):83-98. doi:10.1002/jrsm.1316.
4. Page MJ, Sterne JAC, Higgins JPT, Egger M. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. *Res Synth Methods*. 2021;12(2):248-259. doi:10.1002/jrsm.1468.
5. Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343(jul22 1):d4002-d4002. doi:10.1136/bmj.d4002.